Springer Handbook *of* Science and Technology Indicators

Glänzel
Moed
Schmoch
Thelwall
Editors

Springer

Springer Handbook of Science and Technology Indicators

Springer Handbooks provide a concise compilation of approved key information on methods of research, general principles, and functional relationships in physical and applied sciences. The world's leading experts in the fields of physics and engineering will be assigned by one or several renowned editors to write the chapters comprising each volume. The content is selected by these experts from Springer sources (books, journals, online content) and other systematic and approved recent publications of scientific and technical information.

The volumes are designed to be useful as readable desk book to give a fast and comprehensive overview and easy retrieval of essential reliable key information, including tables, graphs, and bibliographies. References to extensive sources are provided.

Springer Handbook of Science and Technology Indicators

Wolfgang Glänzel, Henk F. Moed,
Ulrich Schmoch, Mike Thelwall (Eds.)

With 279 Figures and 223 Tables

Springer

Editors
Wolfgang Glänzel
ECOOM and Faculty of Economics and Business
KU Leuven
Leuven, Belgium

Henk F. Moed
Amsterdam, The Netherlands

Ulrich Schmoch
Competence Center Policy – Industry – Innovation
Fraunhofer Institute for Systems and Innovation Research ISI
Karlsruhe, Germany

Mike Thelwall
Faculty of Science and Engineering
University of Wolverhampton
Wolverhampton, UK

ISBN 978-3-030-02510-6 e-ISBN 978-3-030-02511-3
https://doi.org/10.1007/978-3-030-02511-3

This Springer imprint is published by the registered company Springer Nature Switzerland AG, part of Springer Nature.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The *Springer Handbook of Science and Technology Indicators* offers a collection of state-of-the-art contributions on quantitative science and technology research. Organized in six parts, the individual chapters focus on various aspects of the development and application of indicators derived from data on scholarly publications, patents, and electronic communication. The 44 chapters are written by leading specialists in the topics selected for this Springer Handbook. These chapters deal with theoretical and methodological issues, illustrate applications, highlight their policy context and relevance, and point to future research directions. In particular, the authors present a survey of the research topics they address, and show their most recent achievements and contribution to the advancement of quantitative studies of science and technology.

The chapters are arranged into six parts:

Part A: Analysis of data sources and network analysis
Part B: Advancement of methodology for research assessment
Part C: Science systems and research policy
Part D: New indicators for research assessment
Part E: Advancement of methodology for patent analysis
Part F: Patent system, patents, and economics.

The Editors' Introduction provides a further specification of the handbook's scope and of the main topics addressed in its chapters. This Springer Handbook aims at four distinct groups of readers: practitioners in the field of science and technology studies; research students in this field; information scientistsa and practitioners in informatics; scientists, scholars, and technicians who are interested in a systematic, thorough analysis of their activities; policy-makers and administrators who wish to be informed about the potentialities and limitations of the various approaches and about their results.

The current handbook can be considered a successor of the *Handbook of Quantitative Science and Technology Studies* edited by Anthony van Raan and published in 1988 and the *Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies of S&T Systems* edited by Henk F. Moed, Wolfgang Glänzel, and Ulrich Schmoch in 2004.

We are grateful to all contributors for their enormous efforts to share their longstanding experience as experts in their research topics and to provide us with excellent chapters for this handbook.

Wolfgang Glänzel
Henk F. Moed
Ulrich Schmoch
Mike Thelwall

Editors' Introduction

The *Springer Handbook of Science and Technology Indicators* continues the tradition and scope set by two predecessor reference works: The *Handbook of Quantitative Studies of Science and Technology*, edited by Anthony F.J. van Raan and published in 1988, and a little more than 15 years later, the *Handbook of Quantitative Science and Technology Research* (editors: Henk Moed, Wolfgang Glänzel, Ulrich Schmoch). Similarly to the previous volumes, this handbook deals with quantitative studies of the science and technology (S&T) system which is conceived as a part of the various national or regional innovation systems.

The current work provides the state of the art of the development and application of methods and models that have been developed to understand and study processes and networks of scientific communication, the indicators for the S&T system that are derived from the documented output of research and patenting activities. Besides reporting and reviewing methodologies and application fields, there is a strong focus on tracing the developments that took place in the field of quantitative S&T studies since the first and second volumes appeared. The three decades since the publication of van Raan's Handbook were characterized by increasing internationalization, the duality of global collaboration and competition in science and technology, challenges to S&T studies that have also created opportunities and proved to be one of the main driving forces for the advancement of our field, and the gradual supplementation and replacement of traditional publishing by electronic communication media and publication channels. Several of these and related issues had already found their way into the second handbook, including the science-technology interface, research collaboration and network analysis, emerging economies and the changing global research landscape, the internationalization of knowledge generation, and data and text mining for S&T studies and webometrics.

The new millennium has sped up development and brought new challenges but also solutions. Increased computing capacity, practically infinite storage capacities, and the development of new algorithms have helped researchers to cope with the challenges of big data that emerged during the last decade. Both the quantity and quality of data now allow the analysis and linkage of huge document corpora, large-scale text mining, and study of the evolution of huge document and actor networks. Open access and open science have improved accessibility to research results and broadened the usage of published information, while scientific blogging provided a platform to communicate science to major stakeholders and the public. The extension of communication and publication channels with new actor and user groups introduced the necessity and possibility of building new measures of usage exchange and networks, which makes it possible to examine new kinds of impact, and to study societal impact beyond the traditional scholarly domain. With the growth of user groups and communities and the wide scope of data sources for information transfer and usage the demand for connectability and interoperability with the necessity of harmonization, standardization, and integration of data emerged (Daraio and Glänzel, 2016).

Software development in recent decades and publicly available data sources, such as Google Scholar and academic licenses for commercial abstract and citation databases with bibliometric features, have opened up bibliometrics to a broader user group among scientists, librarians, and bibliometric semiprofessionals. This has fostered uninformed use of bibliometrics and technometrics, most notably in an evaluative context. This development underlines the necessity of providing an up-to-date handbook on quantitative S&T research to scientists and practitioners, not only reporting the state of the art in the discipline but also giving guidance to practitioners and potential users of S&T indicators.

The contributions to the *Springer Handbook of Science and Technology Indicators* reflect a wide variety of attributes of the contemporary S&T system. Most central concepts have not changed since the previous edition as these include scientific or technological performance, and the productivity or efficiency of the S&T system and its constituent parts. The question of how performance or productivity could be measured also remains a crucial issue but contexts and applications have created new challenges, data provide more and better information, the general trend towards the meso and increasingly micro level has continued, requiring more accuracy and finer granularity. The question of how the various parts in an S&T system react with one another and how this interaction affects the overall performance is still of foremost importance. The need for measuring the impact of research beyond scholarly communication, including policy impact and impact on society, called *broader impact*, has become one of the driving forces for the development of our field. This also includes the necessity of exploiting and mining unstructured and nonstandardized sources for relevant information, a new challenge of big-data processing that has broadened our field further towards computer and information science with the development of new algorithms and new retrieval techniques. These sources and techniques are not replacing the traditional ones based on well-standardized bibliographic and patent databases but they essentially extend traditional quantitative science and technology studies by giving them a new perspective and dimension. To stay abreast of these changes, this handbook includes a new part on the development and application of new indicators (Part D: New Indicators for Research Assessment).

Many new web indicators have been developed since the last handbook was published. In addition to a small set of webometrics, the Altmetrics initiative, started by Jason Priem in 2010, led to the creation of a wide range of new indicators derived from social web sources. It also led to the creation of organizations devoted to collecting and selling altmetric values, including ImpactStory, Altmetric.com, and Plum Analytics. Altmetrics are now widely deployed by publishers in their digital libraries alongside citation counts and are being considered for (limited) research evaluation contexts. Their promise is that they may reflect nonscholarly types of impacts that are ignored by citations from other journal articles, and/or that they may appear more quickly than citations, allowing earlier impact evidence. Most alternative indicators are also easily manipulated and subject to irrelevant uses, however. This makes them difficult to use in most research evaluation contexts. Nevertheless, they can be valuable for self-evaluation and ongoing monitoring purposes (e. g., by funding organizations), if used carefully. They may also be useful on an ad-hoc basis to support narrative claims for nonscholarly impacts. For example, there are instances of this type of use in the case study parts of the 2016 UK Research Excellence Framework evaluations.

In terms of individual new indicators, counts of readers in the social reference manager Mendeley are worth a special mention. These are like citation counts in that they primarily reflect scholarly impact but appear a year earlier, allowing timelier evaluations. In contrast, Altmetric.com scans many policy documents for citations to academic research. Policy mentions give clear evidence of important nonscholarly impacts for academic research. There are also alternative indicators that reflect arts and humanities impacts (e. g., Google Books citations, online reviews), educational impacts (syllabus mentions), and commercial impacts (e. g., Google Patents citations). One of the best-known indicators, tweet counts, appear to primarily reflect publicity, however, and there is no alternative indicator yet that gives good evidence of public interest in research. For this, media mentions are probably the best proxy available. Indicators like download counts are particularly useful for investigating the scholarly communication process by giving insights into patterns of use for documents, including by scholars and students.

In the following, we will give an overview of the chapters that are included in this handbook.

Part A: Analysis of Data Sources and Network Analysis

The chapter by *Vincent Larivière* and *Cassidy R. Sugimoto* deals with the journal impact factor, probably the most widely used bibliometric construct. They argue that this indicator is, by far, the most discussed bibliometric indicator, and has been widely dissected and debated by scholars of every disciplinary orientation. Their chapter presents a brief history of the indicator and highlights a series of well-known limitations, and draws on the existing literature as well as on original research. They highlight the adverse effects of the use of this indicator on authors and publishers, and describe alternative journal-based indicators. Their chapter ends with a call for responsible application of journal indicators, and a commentary on future developments in journal indicators.

Subject delineation has become one of the prominent tasks in bibliometric domain studies. *Michel Zitt*, *Alain Lelu*, *Martine Cadot*, and *Guillaume Cabanac* describe this complex task using three models as a question of disciplines versus invisible colleges. The models, which can be favorably combined with each other, are existing classification schemes, information retrieval, and clustering exercises. The authors discuss the opportunities and limitations of the bibliometric techniques underlying information retrieval, data analysis, and network analysis. They show how multiple network approaches allow the comparison and combination of bibliometric networks. The authors focus on textual and citation networks, but outline possibilities and methods for hybridization. The chapter concludes with a discussion of typical subject delineation schemes and protocols.

In their chapter, *Ronald Rousseau*, *Lin Zhang*, and *Xiaojun Hu* provide a systematic review of interdisciplinarity in scientific research. First, they examine the phenomenon of interdisciplinary research (IDR) from a conceptual perspective and discuss its characteristics and driving forces. The second part is devoted to its quantification and measurement from the information science perspective. The authors proceed from the assumption that IDR is mirrored in the published research documents by the integration of knowledge from different subjects. The authors review existing approaches to measure knowledge integration and show their limitations. Proceeding from the notion by Stirling, Rafols, and Meyer proposing three main components of interdisciplinarity (diversity, variety, and evenness), they propose a true diversity measure. An example from the field of synthetic biology provides an illustration and the chapter concludes with suggestions for future research.

Emilio Delgado López-Cózar, *Enrique Orduna-Malea*, and *Alberto Martín-Martín* argue that the launch of Google Scholar (GS) marked the beginning of a revolution in the scientific information market, because of its automatic indexing of information directly obtained from the web, its ease of use, and its wide coverage. Their chapter lays the foundations for the use of GS as a supplementary source for scientific evaluation, by giving an overview of how GS works, and providing statistics about its size, coverage, and growth rate. In addition, the authors conduct a systematic analysis of the main limitations of GS as an evaluation tool, and compare GS with traditional citation indexes. They conclude that GS presents a broader view of the academic world than the traditional indexes as it includes many previously invisible sources.

The next chapter is devoted to the analysis of current publication trends in gold Open Access (OA). In the first part, *Daniel Torres-Salinas, Nicolas Robinson-Garcia*, and *Henk F. Moed* give a comprehensive review of the current literature on Open Access, specifically in relation to its "citation advantage." This chapter has three dimensions: countries, subject fields, and journals. In the light of this, the authors investigate national gold OA publication patterns, OA journal characteristics and citation differences between gold OA and nongold OA publications, and factors that may affect differences in citation impact between OA and conventional, toll-access journals. The authors also discuss scientists' OA publication strategies and patterns as well as the role of predatory OA journals.

Forecasting future developments in science, technology, and innovation (STI) is the subject of the chapter by *Katy Börner* and *Staša Milojević*. Such forecasts are based on advanced mathematical-statistical and computational models of the STI system, and are facilitated by advances in computational power and in the availability of numerous "big" datasets containing not only bibliometric, but also funding, stock market, social media, and other types of data. Advanced models can be used to simulate and understand the structure and dynamics of the STI system, and enhance human decision-making.

Science mapping in the form of studies of structural aspects of document and actor networks plays an important role in quantitative science studies. The following three chapters tackle this important topic within contemporary scientometric research. The first gives an overview of the advanced bibliometric tool for structural analysis and visualization. The second chapter is devoted to the challenges of the analysis of large-scale bibliometric networks and the third deals above all with fundamental methodological questions of science mapping and topic identification.

Science mapping involves the development and application of computational techniques for the visualization, analysis, and modeling of scientific and technological activities. It is an advanced bibliometric tool to analyze and mine scientific output. *Jose A. Moral-Munoz*, *Antonio G. López-Herrera*, *Enrique Herrera-Viedma*, and *Manuel J. Cobo* review six freely available, comprehensive science mapping tools: Bibexcel, CiteSpace II, CitNetExplorer, SciMAT, Sci2 Tool, and VOSviewer. The authors discuss the strengths and limitations of these tools related to data processing, analysis options, and visualization. They argue that each tool has different properties, and the choice of one over another depends on the type of actors to be analyzed and the type of output expected.

Kevin Boyack and *Richard Klavans* point to the new challenges that have emerged since the last *Handbook of Quantitative Science and Technology Research* was published. The recent science mapping evolution has been facilitated by the availability of full text databases, increased computing capacity, and the development of new algorithms. This has allowed mapping technology to transition from the analysis of small networks to large-scale exercises. The focus is on the analysis of large-scale, global bibliometric networks. The authors give a state-of-the-art report and discuss the commonly used data sources and methods from a historical perspective, continuing to the most recent developments. Their own large-scale topic-level model is used to illustrate the analysis of large-scale bibliometric networks and potential applications.

In his chapter, *Bart Thijs* identifies three drivers of scientometric mapping of science: information-technological innovation; improved community detection; and methodological advancements in the field of scientometrics itself. The author shows that scientometric methodologies using citation-link and lexical approaches lagged the development of the first two drivers. He discusses methodological issues related to community detection. The different approaches to the creation of global maps and the possibility of achieving comparable results at higher levels of granularity are contrasted with the fine-grained solutions possible from local mapping.

Part B: Advancement of Methodology for Research Assessment

Anthony F.J. van Raan gives a comprehensive overview of the methodology and application of advanced bibliometric indicators and introduces bibliometrics as a powerful instrument for the study of science. His historical review starts from the beginning of professional bibliometrics and covers the role of citation indexing in the emergence of the discipline of scientometrics. The review discusses how citation indexing revolutionized quantitative science studies and continues until the stage of contemporary bibliometrics in the internet age. This introduction is followed by a description of advanced state-of-the-art bibliometrics with its rationale and practical needs. The author proceeds from the two main pillars, citation analysis and mapping of science, which can be reduced to a single principle. The author deduces a set of main indicators to be used for research performance assessment with regard to the developments at the Leiden Institute. The conceptual-methodological part is followed by applications of indicators in an evaluative context with various real-life examples. In this context, the author discusses also problematic and controversial issues, such as the use of journal impact factors, the h-index, publication assignment, subject delineation, and university rankings. The last part of the chapter deals with the above-mentioned second pillar of bibliometrics, the mapping of science. Hybrid techniques, the combination of citation analysis and science mapping, and new fields of application are described and discussed.

Ludo Waltman and *Nees Jan van Eck* present a comprehensive overview of a class of bibliometric indicators that are among the most important in bibliometrics, namely field-normalized indicators. The term field indicates a branch of knowledge, such as a research discipline, specialty, or topic. Field-normalized indicators make corrections for differences among fields, so that groups of researchers from different fields can be compared with one another. The authors give an overview of the various field-normalization approaches. Most importantly, they also illustrate how indicators themselves can be evaluated, and how the choice of an approach may affect the outcomes of a bibliometric analysis.

The h-index and its derivatives have become perhaps the most popular and most commonly used bibliometric indicators besides the journal impact factor. Research and applications of Hirsch-type indexes have consequently yielded a large body of literature within our field over the last decade. *András Schubert* and *Gábor Schubert* provide a guided bibliometric tour through more than 3000 papers on this topic. Special attention is paid to the theoretical, mathematical and axiomatic background and various applications as well as the possibility of applying the h-index as a network indicator.

The method of Characteristic Scores and Scales (CSS) was originally proposed in the second half of the 1980s, when their large-scale calculations were still a computational challenge. Because of increased data availability and computational capacity, the method has now become practical. *Wolfgang Glänzel, Bart Thijs*, and *Koenraad Debackere* provide an overview of the various fields of application of this method, which aims to replace the traditional linear approach to citation impact evaluation by a distributional one with a focus on the high end of performance. A discussion of the mathematical background and statistical properties is followed by the implementation of the method in assessment exercises at different levels of aggregation as well as in various disciplinary and multidisciplinary contexts.

The development and application of bibliometric indicators of research performance at the level of individual authors is one of the most debated and complex issues in quantitative science and technology studies. *Lorna Wildgaard* presents a critical overview of the development of this type of indicator. She discusses characteristics and mathematical properties of 68 author-level indicators, and highlights their potential and limitations. The major theme of her contribution is setting the argument for the need to monitor and evaluate current indicator production.

Policy implementation of relevant science, technology, and innovation indicators requires appropriate data management methods, and data integration has become a central issue in this regard. Two main approaches to data integration are in use: procedural and declarative. *Maurizio Lenzerini* and *Cinzia Daraio* follow the latter approach by focusing on the ontology-based data integration (OBDI) paradigm. They discuss the five main principles of this paradigm and the challenges of data integration. Finally, *Sapientia* (the ontology of multidimensional research assessment and its OBDI system) developed at Sapienza University of Rome is provided as an example of an open and collaborative platform for research assessment.

Synergy in innovation systems is studied by *Loet Leydesdorff*, *Inga Ivanova*, and *Martin Meyer* within the framework of the Triple Helix model of university–industry–government relations. This is used as a metaphor in modeling the knowledge-based economy and innovation. Synergy is introduced and analyzed here in the context of the generation of redundancy, the measures of which are derived from an information-theoretic model. Using examples from several countries, it is shown how the Triple-Helix synergy indicator can be applied to analyze regions or sectors in which uncertainty has been significantly reduced and which contribute most to the generation of redundancy. The model and its indicators thus allow the quantification and measurement of the quality of innovation systems at different geographical scales and in terms of sectors.

Part C: Science Systems and Research Policy

The interrelationship between scientometrics and research policy is studied and discussed by *Koenraad Debackere*, *Wolfgang Glänzel*, and *Bart Thijs*. Scientometrics is shown to be a discipline that emerged from the library and information needs of scientific communities and grown into a powerful instrument providing advanced tools and indicators for policy-relevant research assessment. This development is depicted as a symbiosis between scientometrics and science policy. The authors use the example of the *Flemish Expertise Center for R&D Monitoring* (ECOOM) to illustrate this coevolution, pointing to its opportunities, challenges, and limitations.

Research assessment exercises monitoring and evaluating national or regional research performance have a high priority in research management and national research policies. In their chapter, *Sybille Hinze*, *Linda Butler*, *Paul Donner*, and *Ian McAllister* use bibliometric tools to analyze and compare the effects and efficiency of the research assessment regimes of three selected countries (UK, Australia, and Germany). Although the assessment systems of the three countries differ considerably, large differences could not be found regarding their effects and efficiency. They conclude that the systems make less difference than the implementation of an assessment exercise. They further conclude that to understand the mechanisms behind changing performance, indicators are not enough and need to be supplemented by contextual information at various levels of aggregation.

The globalization of research and the use of bibliometric indicators to study this process are the subject of a chapter by *Jacqueline Leta*, *Raymundo das Neves Machado*, and *Roberto Mario Lovón Canchumani*. Given the growing importance of the BRICS countries Brazil, Russia, India, China, and South Africa in the global economy and the science system, the authors focus on scientific collaboration among these countries. They also illustrate how bibliometric techniques can be used to examine traces of the effects of the foundation of the BRICS group upon the international collaboration among its members. A series of techniques was used, including diachronic analysis, Bradford's law, and journal cocitation analysis.

As China publishes over 5000 scientific-scholarly journals, it has developed extensive expertise in journal publishing and journal evaluation. *Zheng Ma* reviews the development of the Chinese journal system in scientific, technical, and medical (STM) fields. The author describes the characteristics of evaluation systems of national journals as compared to those related to international periodicals, in terms of their respective evaluation purposes, evaluation methods, key features, and evaluation criteria. Two cases are presented of China's research work on the evaluation of STM journals, namely the development of the so-called boom index and of comprehensive performance scores for Chinese STM journals. The author also presents analyses of the English-language STM journals in China, and introduces an atomic structure model for evaluating English-language scientific journals published in non-English countries.

Gali Halevi focuses on a crucial issue in science policy, namely the gaps between men and women in the domain of science and scholarship. She provides a thorough review of the various approaches combining bibliometric and other types of research information to the identification of gender among authors of scientific-scholarly literature, and to the measurement of gender disparities. She discusses a series of studies explaining barriers to female participation, and argues that for a comprehensive picture of the underrepresentation of women, bibliometric studies have become an essential tool for tracking not only research participation itself, but also its impact on scientific discovery.

Two chapters are devoted to the measurement of research impact beyond scholarly communication. The first chapter shows how the medical literature is used by clinicians and by the public, while the second reviews and discusses societal impact indicators in recent literature.

The study presented by *Elena Pallari* and *Grant Lewison* analyses how biomedical research could influence its two main goals in improving healthcare: better patient treatment and prevention of illness. They examine two approaches: the research base underlying clinical practice guidelines (CPGs) linked to patient treatment, and stories in the mass media as an expression of healthcare policy. The authors collected CPGs and newsletters from 21 and 22 European countries, respectively, and used Web of Science (WoS) journal articles as their evidence base. The medical research stories from newspapers were linked to research by the WoS papers they cited. The authors found a discrepancy between the papers cited by CPGs and in newspaper stories, on one hand, and those that are frequently cited in scholarly literature, on the other hand. They found that even relatively neglected subject areas could be an important source for medical practice and the general public.

Lutz Bornmann and *Robin Haunschild* give an overview of the literature on societal impact measurement. They first delineate the concept of societal impact, describe the reasons for its emergence, and point to the problems in measuring this kind of impact. Using examples of major projects, they illustrate how frameworks for the measurement of societal impact can be integrated into evaluative contexts. In the last part of their chapter, the authors discuss the possibility of alternative metrics (altmetrics) to measure societal impact.

The use of econometric approaches for the measurement of research productivity, an important concept in research policy and for the wider public, is the subject of a chapter by *Cinzia Daraio*. It explains the benefits of econometric models in research assessment and shows their added value compared to more traditional bibliometric or informetric approaches. Moreover, it gives a theoretical discussion of the nature as well as the ambiguities of the concept of productivity and other key notions in research performance measurement. On the practical side, it presents a checklist for developing econometric models of research assessment.

Gunnar Sivertsen describes the development of a new type of data source for science studies. Institutional and national current research information systems (CRIS) are used to standardize and facilitate research output reporting and research administration. With their high standard of coverage, quality, and standardization, CRIS systems also have the potential to be used as data sources for science studies. Basic requirements are interoperability and data integration at the institutional and national levels. The chapter focuses on challenges and solutions to the development of internationally integrated CRIS. Challenges and possible solutions reaching far beyond the technical are described from the international level to a concrete national example. The authors also show that internationally integrated CRIS can be used for science studies.

Part D: New Indicators for Research Assessment

Indicators for academic outputs derived from social media, such as Twitter, are sometimes known as altmetrics. These are typically quicker to appear than citation counts but are not subject to peer review. These properties make social media indicators fundamentally different from citation-based indicators and there is uncertainty about how they should be used. *Paul Wouters*, *Zohreh Zahedi*, and *Rodrigo Costas* propose principles and conceptual frameworks for using social media data effectively and responsibly in research evaluation contexts. Their chapter gives practical advice as well as theory-based arguments and applies to current as well as future social media indicators.

Monographs, edited books, and book chapters are central to areas of the social sciences and humanities, and can sometimes be important outputs for other researchers. *Alesia Zuccala* and *Nicolas Robinson Garcia* review studies assessing the value of scholarly books. They show that important contributions have been made by four different expert communities, which they define as monitors, subject classifiers, indexers, and indicator constructionists. This unique perspective helps to clarify the advances that need to be made if this relatively under-researched area is to mature as a standard part of the bibliometric landscape.

The pioneering microblog site Twitter is widely used by academics to post about academic publications, such as by announcing the journal articles that they are reading. Counts of tweets about academic outputs are often described as altmetrics. *Stefanie Haustein* reviews studies about the value and interpretation of Twitter altmetrics in many fields. In addition, she uses an analysis of 24 million tweets about scholarly documents to give a detailed exploration of the context of Twitter altmetrics. This information includes the types of documents tweeted about and different types of tweeting patterns, including the problem of nonhuman tweeters.

The social reference manager Mendeley can be used for evidence of the impact of academic publications by counting the number of Mendeley users that have registered them in their personal libraries. *Ehsan Mohammadi* and *Mike Thelwall* discuss how this information is an indicator of the scholarly readership of scholarly outputs and gives citation-like impact evidence. They argue that Mendeley provides earlier impact evidence than citations and its readership data is therefore useful for research evaluations where timely impact data is important. Mendeley is also useful for the background information that it gives about readers, including their job, subject area, and national base.

Any empirical use of web indicators involves gathering data from the web at some stage. This is not as straightforward as downloading citation data from bibliometric or patent databases. *Judit Bar-Ilan* gives a historical overview of methods to gather informetric data from the web, including the main problems and proposed solutions. She shows that researchers have often had to use imperfect methods, such as queries in commercial search engines, to gather their data and this can give misleading results. She emphases that it is important for those collecting data to devote time to data cleansing and other techniques that will help to produce the most accurate and reliable information.

Although scholarly indicators derived from the social web have attracted more attention than indicators derived from the web, the latter seem to be more effective at providing evidence of nonscholarly impacts. *Kayvan Kousha* demonstrates that online data gathering, often through clever standardized queries in commercial search engines, can give indicators of educational, health, informational, general, and other impacts. The methods sometimes take advantage of individual important websites, such as Wikipedia, and sometimes search for evidence from a large part of the web. Kousha's chapter discusses the methods used to generate a range of web-based indicators and reviews evidence of their limitations and value.

Scholarly articles are usually available in electronic form and sometimes only in digital versions. Before an article can be cited, it must be accessed and this may well involve downloading it from a publisher website or a digital repository. Data from such sources may therefore give earlier evidence of the academic impact of publications and perhaps also evidence of interest from nonpublishing audiences. *Edwin A. Henneken* and *Michael J. Kurtz* demonstrate how analyzing the log files of a digital repository can give new types of detailed information about how academic research is accessed. They illustrate this with a detailed analysis of clickstream data from the Astrophysics Data System of peer reviewed and other publications about astronomy and physics.

Although most research evaluations focus mainly on journal articles and perhaps also books, a range of other types of activity and output are important to the missions of academia. It is important that the contributions of scholars producing nonstandard outputs are recognized and one way of achieving this is by generating impact indicators to support qualitative claims for their value or impact. *Mike Thelwall* introduces online indicators for different types of scholarly output that take advantage of easily available online quantitative data, such as view or download counts published online or available for automatic harvesting. Whilst interpreting the numerical information is complex due to the variety of different goals and audiences for ostensibly similar outputs, the survey shows that useful data is often available, although always with limitations.

Part E: Advancement of Methodology for Patent Analysis

The present handbook covers both science and technology indicators, as they are closely intertwined. Science is primarily linked to publication indicators, and technology to patent indicators. Nevertheless, the logic behind publication and patent statistics are quite different, so that multiple chapters are needed to explain different aspects of patent statistics. Furthermore, in addition to patents, trademarks and standards are also used as technology indicators. These alternative indicators are discussed in two chapters.

A general challenge of patent analysis is the growing number of patent applications, so that it is increasingly difficult to identify all relevant documents referring to a specific topic. For this purpose, *Carson K. Leung*, *Wookey Lee*, and *Justin Jongsu Song* present an advanced text-based retrieval system and compare three different retrieval algorithms. With their approach, patent documents that are relevant to keyword terms in a user query can be retrieved efficiently without returning many irrelevant patent documents.

A specific characteristic of patents in comparison to publications is that the text is not written by the inventors, but by patent attorneys. The latter have a decisive influence on the successful process of a patent application. *Rainer Frietsch* and *Peter Neuhäusler* analyze the differences between experienced and less experienced attorneys in more detail.

A further peculiarity of patents is that most include images of the invention. On this basis, images can be used for patent retrieval as well. *Ilias Gialampoukidis*, *Anastasia Moumtzidou*, *Stefanos Vrochidis*, and *Ioannis Kompatsiaris* describe different computer-based approaches and illustrate them with examples.

Ulrich Schmoch and *Mosahid Khan* deal with new methodological issues of retrieval for patent indicators linked to the change of the patent system in the last 20 years and the new ways to access patent data. It describes international flows of patent applications between the US, Europe, and Southeast Asia and illustrates methods for an appropriate cross-country comparison. A central topic of this chapter is the implications of the frequently used Patent Cooperation Treaty (PCT) route of patent applications on the conception of search strategies and the interpretation of search results. Furthermore, the possibilities of search with the new international Cooperative Patent Classification (CPC) are explained. In addition, the patenting activities of very large companies and the patent value are discussed.

For knowledge-based technologies, scientific and technological activities are performed in parallel and influence each other. Here it is useful to identify similar patents and publications. *Tom Magerman* and *Bart Van Looy* present linguistic text mining approaches to identify similarities and illustrate them with examples. They discuss the advantages and disadvantages of different retrieval methods.

Contributions in this handbook show that more powerful computer systems have increased the power of text mining in science and technology analysis. *Samira Ranaei*, *Arho Suominen*, *Alan Porter*, and *Tuomo Kässi* give a broad literature review of the most relevant approaches and show by examples their usefulness. For instance, text mining classification analysis of patents can lead to additional results. It is insightful to compare the methods of other chapters on text mining in this handbook to the assessment of the authors of this chapter.

Other contributions in this handbook also illustrate the potential of text mining. However, the quality of results are influenced by the quality of the texts analyzed. Thus, the yield of text-based retrieval at some patent offices is higher than at others depending on the legal requirements for technical disclosure of the patent abstracts. For instance, the text quality at the US Patent and Trademark Office is commendable.

An advantage of patents compared to scientific publications is their detailed classifications, so that in many cases a precise definition of a topic is feasible. However, it is possible that similar items are classified in different parts of the classification, so that it is difficult to identify all documents relevant to a topic. *Andrea Bonaccorsi*, *Gualtiero Fantoni*, *Riccardo Apreda*, and *Donata Gabelloni* suggest a functional classification system for patents which supports new types of patent searches based on functional dictionaries. Again, the approach is based on advanced text mining. The authors present some examples of contexts for which their approach is useful.

Part F: Patent System, Patents, and Economics

A widespread misconception is that software inventions can be patented only in the USA, but not in Europe. The chapter of *Peter Neuhäusler* and *Rainer Frietsch* shows that in many cases software can also be patented in Europe and that the share within all patent applications is steadily increasing. In addition, the introduction of the subclass G06Q (data processing systems or methods, specially adapted for administrative, commercial, financial, managerial, supervisory, or forecasting purposes) in 2006 confirms that the attitude of the European Patent Office about software has substantially changed.

Patents and trademarks are competitive tools for research-intensive technologies. *Sandro Mendonça*, *Ulrich Schmoch*, and *Peter Neuhäusler* show with the example of enterprises from the EU Industrial R&D Scoreboard that trademarks, product and service marks have become increasingly important, especially in the case of service marks. They illustrate with examples the fact that the strategies of enterprises for patents and trademarks vary considerably between sectors and sometimes even within sectors. They argue that trademarks should be considered in parallel to patents whenever possible.

In most countries, the annual number of patent applications is stable and changes only in the long term. A new phenomenon for threshold countries, such as South Korea and China, is the tremendous increase in the annual number of patent applications within a decade. *Chan-Yuan Wong* and *Hon-Ngen Fung* analyze South Korea and China, highlighting parallel increases in scientific activities. These types of explorations will be important for understanding the worldwide landscape of science and technology and the emergence of a new regime in international trade.

The final contribution to the technology section of this handbook is concerned with standards. It shows that new technologies are not sufficient for market success, but that the development of standards is a further decisive step. *Knut Blind* explains why standards are important for technology and that they can be used as indicators for describing supplementary aspects of technological performance.

References

- A. F. J. van Raan (Ed.): *Handbook of Quantitative Studies of Science and Technology* (North Holland, Amsterdam 1988)
- H. Moed, W. Glänzel, U. Schmoch (Eds.): *Handbook of Quantitative Science and Technology Research* (Springer, Dordrecht 2004)
- C. Daraio, W. Glänzel: Grand challenges in data integration—state of the art and future perspectives: An introduction, Scientometrics **108**(1), 391–400 (2016)

Wolfgang Glänzel
Henk F. Moed
Ulrich Schmoch
Mike Thelwall

About the Editors

Wolfgang Glänzel is Director of the Center for Research and Development Monitoring (ECOOM) and Full Professor at KU Leuven, Belgium. He is also affiliated with the Department of Science Policy and Scientometrics at the Library of the Hungarian Academy of Sciences in Budapest, Hungary. Wolfgang Glänzel studied Mathematics at the Eötvös Lorand University (ELTE) in Budapest. He holds a doctorate in Mathematics from ELTE obtained in 1984 as well as a PhD in the Social Sciences which he obtained from Leiden University, The Netherlands, in 1997. He is also Guest Professor at several Universities in China and the UK.

Wolfgang Glänzel's research activities comprise probability theory and mathematical statistics, quantitative science studies, and research policy. In the field of probability theory he has published on the characterization of probability distributions, in quantitative science studies and research policy his research is mainly focused on models of scientific communication, the development of scientometric indicators, the analysis of coauthorship patterns and research collaboration, the structural-cognitive mapping of science, bibliometrics-aided retrieval, and the measurement of research performance at various levels of aggregation. Wolfgang Glänzel has published numerous journal articles, proceedings, and book chapters and coauthored/coedited several books. He was Alexander von Humboldt Fellow for two years in Germany. He has published in probability theory and mathematical statistics, computer science and various topics in quantitative science studies, scientometrics, and informetrics.

In 1999, Professor Glänzel received the international *Derek deSolla Price Award* for outstanding contributions to the quantitative studies of science. He is Secretary-Treasurer of the *International Society for Scientometrics and Informetrics* (ISSI) and Editor-in-Chief of the Society's Newsletter. Wolfgang Glänzel has been Editor-in-Chief of the international journal *Scientometrics* since 2014 and Academic Editor of the multidisciplinary OA journal PLoS ONE since 2013.

Henk F. Moed holds a PhD degree in Science Studies and was a senior staff member and Full Professor in Research Assessment Methodologies at the Centre for Science and Technology Studies at Leiden University, The Netherlands between 1981 and 2010. He was a Senior Scientific Advisor at Elsevier, Amsterdam, from 2010–2014.

Henk Moed has been active in numerous research topics, including: the creation of bibliometric databases; the development and use of bibliometric indicators; studies of the effects of *Open Access*; bibliometric studies of international scientific migration and collaboration; comparisons of Web of Science, Scopus and Google Scholar; multi-dimensional assessment of research impact; the potential of altmetrics; ontology-based bibliometric data management.

Dr. Moed published over 100 research articles in international, peer-reviewed journals, was Program Chair of numerous international conferences in the field, and is an Editorial Board Member of several journals in his field. He was a winner of the *Derek deSolla Price Award* for outstanding contributions to the quantitative studies of science in 1999. He is currently an independent scientific advisor, and Visiting Professor at the Sapienza University of Rome and University of Granada.

Ulrich Schmoch holds a Doctoral Degree in Social and Political Sciences (University of Hannover, Germany), a Diploma of Mechanical Engineering (University of Hannover) with modules at the University of Grenoble, France, and a State Doctorate in Sociology of Science and Technology (Karlsruhe Institute of Technology KIT, Germany). From 1983–1985, he was employed in the office of a patent attorney.

Since 1986, Professor Schmoch has worked at the Fraunhofer Institute for Systems and Innovation Research (ISI), Karlsruhe, Germany, and performed and coordinated about 200 projects in the areas of innovation indicators, patent analysis, bibliometrics, systems of innovation, technology transfer, science management, innovation in threshold countries, innovations in the service sector, science-based technologies, and organization and mission of universities in an international comparison.

Between 1991–2002, Ulrich Schmoch was Deputy Head and from 2002–2004, Head of the Department *Technology Assessment and Innovation Strategies*. In 2011, he taught innovation indicators and patent statistics at the University of Stellenbosch, South Africa. Between February 2012 and September 2014 he taught as Stand-In Professor in Science Management at the German University of Public Administration in Speyer. Since 2015, he has once again been working at the Fraunhofer ISI as a Senior Researcher.

Mike Thelwall is a Professor of Data Science and leads the Statistical Cybermetrics Research Group at the University of Wolverhampton, UK, which he joined in 1989. He is also Docent at the Department of Information Studies at Åbo Akademi University in Turku, Finland, and a Research Associate at the Oxford Internet Institute. He obtained a PhD in Pure Mathematics from the University of Lancaster, UK.

Professor Thelwall has developed tools and methods for gathering and analyzing webometric and altmetric alternative indicators. His publications include 300 refereed journal articles, 30 book chapters, and 3 books.

Mike Thelwall is an Associate Editor of the Journal of the Association for Information Science and Technology and sits on four other editorial boards. He was a member of the UK's independent review of the role of metrics in research assessment (2014–2015) and is a member of the UK Forum for Responsible Research Metrics.

List of Authors

Riccardo Apreda
Erre Quadro s.r.l.
Pisa, Italy

Judit Bar-Ilan
Information Science
Bar-Ilan University
Ramat Gan, Israel

Knut Blind
Institute of Technology and Management
Technische Universität Berlin
Berlin, Germany

Andrea Bonaccorsi
DESTEC
University of Pisa
Pisa, Italy

Katy Börner
School of Informatics, Computing,
and Engineering
Indiana University
Bloomington, IN, USA

Lutz Bornmann
Division for Science and Innovation Studies
Administrative Headquarters of the
Max Planck Society
Munich, Germany

Kevin W. Boyack
SciTech Strategies, Inc.
Albuquerque, NM, USA

Linda Butler
School of Politics and International Relations
The Australian National University
Acton, Australia

Guillaume Cabanac
Computer Science Department
University of Toulouse
Toulouse, France

Martine Cadot
Knowledge and Language Management
LORIA (Laboratory in Computer Science and its
Applications)
Villers-lès-Nancy, France

Manuel J. Cobo
Department of Computer Science and Engineering
University of Cádiz
Algeciras, Spain

Rodrigo Costas
Centre for Science and Technology Studies (CWTS)
Leiden University
Leiden, The Netherlands

Cinzia Daraio
Department of Computer, Control,
and Management Engineering
Sapienza University of Rome
Rome, Italy

Koenraad Debackere
ECOOM and Faculty of Economics and Business
KU Leuven
Leuven, Belgium

Emilio Delgado López-Cózar
Información y Documentación
Universidad de Granada
Granada, Spain

Paul Donner
Research Area 2, Research System and Science
Dynamics
German Centre for Higher Education Research
and Science Studies, DZHW
Berlin, Germany

Nees Jan van Eck
Centre for Science and Technology Studies (CWTS)
Leiden University
Leiden, The Netherlands

Gualtiero Fantoni
Department of Civil and Industrial Engineering
University of Pisa
Pisa, Italy

Rainer Frietsch
Competence Center Policy – Industry – Innovation
Fraunhofer Institute for Systems and Innovation
Research ISI
Karlsruhe, Germany

Hon-Ngen Fung
Post-graduate and Research Centre
KDU University College
Shah Alam, Malaysia

Donata Gabelloni
Erre Quadro s.r.l.
Pisa, Italy

Ilias Gialampoukidis
Information Technologies Institute
Centre for Research and Technology Hellas
Thermi, Greece

Gali Halevi
Icahn School of Medicine at Mount Sinai
New York, NY, USA

Robin Haunschild
Central Information Service
Max Planck Institute for Solid State Research
Stuttgart, Germany

Stefanie Haustein
School of Information Studies
University of Ottawa
Ottawa, Canada

Edwin A. Henneken
High Energy Physics
Harvard-Smithsonian Center for Astrophysics
Cambridge, MA, USA

Enrique Herrera-Viedma
Department of Computer Science and Artificial Intelligence
University of Granada
Granada, Spain

Sybille Hinze
Research System and Science Dynamics
German Center for Higher Education Research and Science Studies (DZHW)
Berlin, Germany

Xiaojun Hu
School of Basic Medical Sciences
Zhejiang University
Hangzhou, China

Inga Ivanova
Institute for Statistical Studies and Economics of Knowledge
National Research University Higher School of Economics (NRU HSE)
Moscow, Russian Federation

Tuomo Kässi
School of Engineering Science, Industrial Engineering and Management
Lappeenranta University of Technology (LUT)
Lappeenranta, Finland

Mosahid Khan
World Intellectual Property Organization (WIPO)
Geneva, Switzerland

Richard Klavans
SciTech Strategies, Inc.
Wayne, PA, USA

Ioannis Kompatsiaris
Information Technologies Institute
Centre for Research and Technology Hellas
Thermi, Greece

Kayvan Kousha
Statistical Cybermetrics Research Group
University of Wolverhampton
Wolverhampton, UK

Michael J. Kurtz
Optical and Infrared Astronomy
Harvard-Smithsonian Center for Astrophysics
Cambridge, MA, USA

Vincent Larivière
École de bibliothéconomie et des sciences de l'information
Université de Montréal
Montréal, Canada

Wookey Lee
Industrial Engineering
Inha University
Incheon, South Korea

Alain Lelu
ELLIADD (retired)
Université de Franche-Comté
Besançon, France

Maurizio Lenzerini
Department of Computer, Control, and Management Engineering
Sapienza University of Rome
Rome, Italy

Jacqueline Leta
Programa de Educação, Gestão e Difusão / Instituto de Bioquímica Médica
Universidade Federal do Rio de Janeiro
Rio de Janeiro, Brazil

Carson Leung
Department of Computer Science
University of Manitoba
Winnipeg, Canada

Grant Lewison
Division of Cancer Studies
King's College London
London, UK

Loet Leydesdorff
Amsterdam School of Communication Research (ASCoR)
University of Amsterdam
Amsterdam, The Netherlands

Antonio G. López-Herrera
Department of Computer Science and Artificial Intelligence
University of Granada
Granada, Spain

Roberto Mario Lovón Canchumani
Graduate Program in Information Science/IBICT-UFRJ
Federal University of Rio the Janeiro
Rio de Janeiro, Brazil

Zheng Ma
Institute of Scientific and Technical Information of China (ISTIC)
Beijing, China

Tom Magerman
Faculty of Economics and Business
KU Leuven
Leuven, Belgium

Alberto Martín-Martín
Información y Documentación
Universidad de Granada
Granada, Spain

Ian McAllister
School of Politics and International Relations
The Australian National University
Acton, Australia

Sandro Mendonça
ISCTE University Institute of Lisbon
Lisboa, Portugal

Martin Meyer
Kent Business School
University of Kent
Canterbury, UK

Staša Milojević
School of Informatics, Computing, and Engineering
Indiana University
Bloomington, IN, USA

Ehsan Mohammadi
School of Library and Information Science
University of South Carolina
Columbia, SC, USA

Jose A. Moral-Munoz
Department of Nursing and Physiotherapy
University of Cádiz
Cádiz, Spain

Anastasia Moumtzidou
Information Technologies Institute
Centre for Research and Technology Hellas
Thermi, Greece

Peter Neuhäusler
Competence Center Policy – Industry – Innovation
Fraunhofer Institute for Systems and Innovation Research ISI
Karlsruhe, Germany

Raymundo das Neves Machado
Departamento de Fundamentos e Processos Informacionais, Instituto de Ciência da Informação
Universidade Federal da Bahia
Salvador, Brazil

Enrique Orduña-Malea
Audiovisual Communication, Documentation, and History of Art
Universitat Politècnica de València
Valencia, Spain

Elena Pallari
Centre for Implementation Science, Health Service & Population Research Department
King's College London
London, UK

Alan Porter
Search Technology, Inc.
Norcross, GA, USA

Anthony van Raan
Centre for Science and Technology Studies (CWTS)
Leiden University
Leiden, The Netherlands

Samira Ranaei
School of Engineering Science, Industrial Engineering and Management
Lappeenranta University of Technology (LUT)
Lappeenranta, Finland

Nicolas Robinson-García
School of Public Policy
Georgia Institute of Technology
Atlanta, GA, USA

Ronald Rousseau
ECOOM
KU Leuven
Leuven, Belgium

András Schubert
Library and Information Center
Hungarian Academy of Sciences
Budapest, Hungary

Gábor Schubert
University Library
Stockholm University
Stockholm, Sweden

Gunnar Sivertsen
NIFU – Nordic Institute for Studies of Innovation, Research and Education
Oslo, Norway

Justin Jongsu Song
Industrial Engineering
Inha University
Incheon, South Korea

Cassidy R. Sugimoto
School of Informatics and Computing
Indiana University Bloomington
Bloomington, IN, USA

Arho Suominen
VTT Technical Research Centre of Finland
Espoo, Finland

Bart Thijs
ECOOM
KU Leuven
Leuven, Belgium

Daniel Torres-Salinas
EC3metrics SL y Medialab-UGR
Universidad de Granada
Granada, Spain

Bart Van Looy
Faculty of Economics and Business
KU Leuven
Leuven, Belgium

Stefanos Vrochidis
Information Technologies Institute
Centre for Research and Technology Hellas
Thermi, Greece

Ludo Waltman
Centre for Science and Technology Studies (CWTS)
Leiden University
Leiden, The Netherlands

Lorna Wildgaard
Researcher Support
The Royal Library
Copenhagen, Denmark

Chan-Yuan Wong
Institute of Technology Management
National Tsing Hua University
Hsinchu City, Taiwan

Paul Wouters
Centre for Science and Technology Studies (CWTS)
Leiden University
Leiden, The Netherlands

Zohreh Zahedi
Centre for Science and Technology Studies (CWTS)
Leiden University
Leiden, The Netherlands

Lin Zhang
School of Information Management
Wuhan University
Wuhan, China

Michel Zitt
Lereco Lab (retired)
National Institute for Agronomic Research (INRA), Dept SAE2
Nantes, France

Alesia Zuccala
Department of Information Studies
University of Copenhagen
Copenhagen, Denmark

Contents

Part C Science Systems and Research Policy

Part F Patent System, Patents and Economics

List of Abbreviations

5G	fifth-generation wireless system

A

ABM	agent-based model
ADR	accumulation for dynamic ranking
ADS	astrophysics data system
AHDH	adaptive hierarchical density histogram
AHP	analytic hierarchy process
AIA	America Invents Act
AIDS	acquired immune deficiency syndrome
AIS	article influence score
AKM	axial *k*-means
ALIRP	author-level indicators of research production
ALM	article-level metrics
AMP	agent modeling platform
AN	attribute network
ANOVA	analysis of variance
ANT	actor–network theory
AP	advanced placement
AP	average precision
APC	article processing charges
API	application programming interface
ARC	Australian Research Council
ARI	adjusted Rand index
ASJC	All Science Journal Classification
ASM	atomic structure model
ASR	American Sociological Review
ATECO	attività economiche
AVA	autovalutazione, valutazione periodica, accreditamento
A&H	arts and humanities
A&HCI	Arts and Humanities Citation Index

B

BC	bibliographic coupling
BCCL	book classification for Chinese libraries
BDOA	Berlin Declaration on Open Access
Belspo	Belgian Federal Science Policy Office
BIS	bibliographic information system
BKCI	Book Citation Index
BKCI-S	Book Citation Index – Science
BKCI-SSH	Book Citation Index – Social Sciences & Humanities
BM25	best match 25
BMC	BioMed Central
BMJ	British Journal of Medicine
BOAI	Budapest Open Access Initiative
BOF	extraordinary research fund
BoW	bag-of-words
BRICS	Brazil, Russia, India, China and South Africa

C

CA	correspondence analysis
CAGR	compound annual growth rate
CARDI	cardiovascular disease
CAS	chemical abstracts service
CASRAI	Consortia Advancing Standards in Research Administration Information
CBA	cost–benefit analysis
CBIR	content-based image retrieval
CC	co-citation
ccTLD	country code top-level domain
CHSSCD	Chinese Humanities and Social Sciences Core Journals Database
CII	computer-implemented invention
CISE	computer and information science and engineering
CJCR	Chinese STM Citation Report
CJK	China, Japan, South Korea
CNCI	category normalized citation impact
COL	colorectal cancer
COPD	chronic obstructive pulmonary disease
COR	coronary heart disease
CoreSC	core scientific concept
COUNTER	Counting Online Usage of Networked Electronic Resources
CPC	Cooperative Patent Classification
CPCI-S	Conference Proceedings Citation Index – Science
CPCI-SSH	Conference Proceedings Citation Index – Social Science & Humanities
CPG	clinical practice guideline
CPS	comprehensive performance score
CQ	conjunctive queries
CRIS	current research information system
CRISTIN	Current Research Information System in Norway
CSCD	Chinese Science Citation Database
CSS	characteristic scores and scales
CSSCI	Chinese Social Sciences Citation Index
CSTPCD	Chinese Scientific and Technical Papers and Citations Database
CSV	comma-separated values
CWTS	Centre for Science and Technology Studies

D

DALY	disability-adjusted life years
DBSCAN	density-based spatial clustering of application with noise
DC	direct citation
DCI	data citation index
DDC	Dewey Decimal Classification System

DDR	dispersion for dynamic ranking
DEA	data envelopment analysis
DGP	data generating process
DIABE	diabetes
DL	digital library
DOAJ	Directory of Open Access Journals
DOI	digital object identifier
DORA	San Francisco Declaration on Research Assessment
DORA	Declaration on Research Assessment
DSM	distributional semantic model
DSSC	dye-sensitized solar cells

E

ECLA	European classification
EDC	extended direct citation
EEA	European Economic Area
Ei	Engineering index
EM	expectation maximization algorithm
EMNPC	equalized mean-based normalized proportion cited
EMR	Elastic MapReduce
EOAC	edge orientation autocorrelogram
EPC	European Patent Convention
ER	entity resolution
ERA	Excellence in Research for Australia
ERiC	Evaluating Research in Context
ERIH	European Reference Index for the Humanities
ESCI	Emerging Sources Citation Index
ESI	Essential Science Indicators
ET	emerging technology
ETER	European Tertiary Education Register
EU FP7/H2020	EU Framework Program/Horizon 2020
EuroCRIS	European Current Research Information Systems
EV	electric vehicle

F

FAIR	findability, accessibility, interoperability, and reusability
FBS	function–behavior–structure
FCEV	fuel cell electric vehicle
FCM	fuzzy C-means method
FDI	foreign direct investment
FECR	field expected citation rate
FN	false negative
FoR	fields of research
FoS	fields of science and technology
FP	false positive
FPC	functional patent classification
FRBR	Functional Requirements for Bibliographic Records
FSS	fractional scientific strength
FTE	full time equivalent
FUSE	Foresight and Understanding from Scientific Exposition
FWCI	field-weighted citation impact
FWO	fund for scientific research Flanders

G

GAV	global-as-view
GDP	gross domestic product
GIF	Global Impact Factor
GLM	generalized linear model
GML	Graph Modeling Language
GMM	generalized method of moment
GPL	general public license
GPU	graphics processing unit
GRID	Global Research Identifier Database
GS	Google Scholar
GSC	Google Scholar Citations
GSM	Google Scholar Metrics
GtR	Gateway to Research
GUESS	Graph Exploration system
GUI	graphical user interface

H

HCE	hit count estimate
HCR	highly cited researchers
HHI	Herfindahl–Hirschman index
HIV	human immunodeficiency virus
HTML	hypertext markup language
HWWS	handwashing with soap

I

ICA	independent component analysis
ICE	internal combustion engine
ICEE	Indicator of Quality for Publishers according to Experts
ICT	information and communication technology
ICV	Index Copernicus metric value
IDF	inverted document frequency
IDR	interdisciplinary research
IF	impact factor
IFBSCP	Impact Factor Biased Self-Citation Practices
IM	intermediary
INPADOC	International Patent Documentation
IoT	Internet of Things
IP	intellectual property
IPC	International Patent Classification
IPR	intellectual property rights
IR	information retrieval
ISBN	International Standard Book Number
ISI	Institute for Scientific Information
ISNI	International Standard Name Identifier
ISO	International Organization for Standardization
IT	information technology

J

J-STAGE	Japan Science and Technology Information Aggregator
JASIST	Journal of the Association for Information Science and Technology
JASSS	Journal of Artificial Societies and Social Simulation
JCR	Journal Citation Reports
JID	Journal IDentification
JIF	Journal Impact Factor

K

KET	key enabling technology
KIS	knowledge-intensive services
KML	keyhole markup language
KTN	knowledge transfer network
KWD	keyword weight distribution

L

LBC	Library-Bibliographical Classification
LBD	literature-based discovery
LBT	link-bridged topic model
LC	longitudinal coupling
LCC	Library of Congress Classification
LDA	latent Dirichlet allocation
LIS	library and information science
LMI	lead market initiative
LOWESS	locally weighted scatterplot smoothing
LSA	latent semantic analysis
LSH	locality-sensitive hashing
LSI	latent semantic indexing
LUBM	Lehigh University Benchmark

M

MA	Microsoft Academic
MAG	Microsoft Academic Graph
MBS	mean blog score
MCS	mean citation score
MDS	multidimensional scaling
MECR	mean expected citation rate
MENTH	mental disorder
MeSH	Medical Subject Headings
MHT	medium-high-technology
MNCS	mean normalized citation score
MNE	multinational enterprise
MNLCS	Mean Normalized Log-transformed Citation Score
MOCR	mean observed citation rate
MRI	magentic resonance imaging
MRS	mean readership score
MSC	mathematics subject classification
MTS	mean Twitter score
M&A	mergers and acquisitions

N

NACE	Statistical Classification of Economic Activities in the European Community
NB	naive Bayes
NBIC	nanotechnology, biotechnology, information technology, and cognitive science
NCD	normal compression distance
NEST	newly emerging science and technology
NGD	normalized Google distance
NGO	nongovernmental organization
NLP	natural language processing
NLTK	Natural Language Toolkit
NMF	non-negative matrix factorization
NMI	normalized mutual information
NPD	new product development
NPR	nonpatent reference
NVL	Nationale VersorgungsLeitlinien
nwD	neighbor-weighted degree
NWE	neural word embeddings

O

OA	open access
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OAMJ	open access mega-journal
OBDA	ontology-based data access
OBDI	ontology-based data integration
OBDM	ontology-based data management
OCR	optical character recognition
OCR	over-citation ratio
OJS	open journal systems
OLS	ordinary least squares
OOB	out-of-bag
OP	open factor
ORCID	Open Researcher and Contributor ID
OWL	ontology web language

P

PACS	physics and astronomy classification scheme
PAN	publication-attribute network
PATSTAT	EPO Worldwide Patent Statistical Database
PCA	principal component analysis
PCT	Patent Cooperation Treaty
PEM	proton-exchange membrane
PFI	Pact for Research and Innovation
PHI	PartnersHIp Ability Index
PI	principal investigator
PISA	Program for International Student Assessment
PLoS	Public Library of Science
pLSA	probabilistic latent semantic analysis
PLSI	probabilistic latent semantic indexing
PN	publication network

PNG	portable network graphics
POS	part of speech
PP(top 10%)	proportion of top 10% publications
ppp	purchasing power parity
PPV	positive predictive value
PRI	public research institution
PRO	public research organization

Q

QMS	quality management system

R

RAE	Research Assessment Exercise
RCD	Research Core Data Set
RCI	relative citation impact
RDBMS	relational database management system
RDI	research and development intensity
REF	Research Excellence Framework
RePEc	Research Papers in Economics
REPP	research embedment and performance profile
RESSH	Research Evaluation in the Social Sciences and Humanities
RF	random forest
RG	ResearchGate
RIS	research information system
RJ	Riksbankens Jubileumsfond
RO	receiving office
RoW	rest of the world
RPYS	referenced publication years spectroscopy
RQ	research quantum
RQF	Research Quality Framework
RSI	relative specialization index
R&D	research and development
R&I	research and innovation

S

SAO	subject–action–object
SAS	Statistical Analysis System
SBIR	Small Business Innovation Research
SC	subject category
SCI	Science Citation Index
Sci^2	Science of Science
SCI-WoS	Science Citation Index of the Web of Science
SCIE	Science Citation Index Expanded
SciMAT	science mapping analysis software tool
SciSTIP	Scientometrics and Science, Technology and Innovation Policy
SERP	search engine results page
SFA	stochastic frontier analysis
SIAMPI	Social Impact Assessment Methods for research and funding instruments
SIFT	scale invariant features transform
SJR	SCImago Journal Rank
SLM	smart local moving
SLMA	smart local moving algorithm
SMART	specific, measurable, accessible, relevant, and traceable
SME	small and medium-sized enterprise
SMOTE	synthetic minority oversampling technique
SNA	system of national account
SNIP	source normalized impact per paper
SNLG	Sistema Nazionale per le Linee Guida
SOM	self-organizing map
SOOS	steunpunt o&o statistieken
SOSP	Science of Science Policy
SPSS	statistical package for the social sciences
SQL	structured query language
SS	social sciences
SSCI	Social Sciences Citation Index
SSH	social sciences and humanities
ST	single terms
ST	science and technology
STEM	science, technology, engineering, and mathematics
STI	science, technology, and innovation
STICCI	software tool for improving and converting citation indices
STIP	science, technology, and innovation policy
STM	scientific, technical and medical
STN	Science Technology Network
STS	SciTech Strategies
STS	science and technology studies
SURF	speeded-up robust features
SVD	singular value decomposition
SVG	scalable vector graphics
SVM	support vector machine
S&T	science and technology

T

TA	threshold algorithm
TARL	topics, aging, and recursive linking
tAS	triple adjacent segment
TBT	technical barrier to trade
TED	technology, entertainment, design
TF	term frequency
TF	technological forecasting
TF-ICF	term frequency-inverse corpus frequency
tf-idf	term frequency-inverse document frequency
TF-IPCF	term frequency-inverse patent category frequency
TH	triple helix
THE	Times Higher Education
TIMSS	Trends in International Mathematics and Science Study

TLD	top-level domain
TN	true negative
TOA	technological opportunities analysis
TOPSIS	technique for order of preference by similarity to ideal solution
TP	true positive
TPR	true positive rate
TRIPS	Trade Related Aspects of Intellectual Property Rights
TRIZ	theory of inventive problem solving
TRM	technology road mapping
TTIP	Transatlantic Trade and Investment Partnership

U

UDC	universal decimal classification
UIF	Universal Impact Factor
UML	unified modeling language
UNA	unique name assumption
UoA	unit of assessment
URL	uniform resource locator
USPC	US patent classification

V

VABB-SHW	Vlaams Academisch Bibliografisch Bestand voor de Sociale en Humane Wetenschappen
VBA	visual basic application
VEM	variational expectation maximization
VQR	Valutazione della Qualità della Ricerca—Evaluation of Research Quality
VSM	vector space model

W

WIF	web impact factor
WIRe	web impact report
WoS	Web of Science
WPI	World Patents Index
WSDM	Web Search & Data Mining

X

XML	extensible markup language
XRAC	cross-reference art collection

Part A

Analysis

Part A Analysis of Data Sources and Network Analysis

1. The Journal Impact Factor: A Brief History, Critique, and Discussion of Adverse Effects

Vincent Larivière, Cassidy R. Sugimoto

The journal impact factor (JIF) is, by far, the most discussed bibliometric indicator. Since its introduction over 40 years ago, it has had enormous effects on the scientific ecosystem: transforming the publishing industry, shaping hiring practices and the allocation of resources, and, as a result, reorienting the research activities and dissemination practices of scholars. Given both the ubiquity and impact of the indicator, the JIF has been widely dissected and debated by scholars of every disciplinary orientation. Drawing on the existing literature as well as original research, this chapter provides a brief history of the indicator and highlights well-known limitations—such as the asymmetry between the numerator and the denominator, differences across disciplines, the insufficient citation window, and the skewness of the underlying citation distributions. The inflation of the JIF and the weakening predictive power is discussed, as well as the adverse effects on the behaviors of individual actors and the research enterprise. Alternative journal-based indicators are described and the chapter concludes with a call for responsible application and a commentary on future developments in journal indicators.

1.1 Origins of the Journal Impact Factor

In the 1975 version of the Science Citation Index (SCI), Eugene Garfield and the Institute for Scientific Information (ISI) added a new component to their information products: the Journal Citation Reports (JCR). While *Garfield* and *Sher* proposed the concept of an *impact factor* as early as 1963 [1.1]—and tested it at a larger scale in 1972 [1.2]—the 1975 JCR was ISI's first comprehensive reporting of their data at the journal level. On the basis of more than 4.2 million references made in 1974 by 400 000 papers published in about 2400 journals, this new information source provided a detailed list of journal-to-journal citation linkages, as well as the first iteration of what would become the most discussed and derided bibliometric indicator: the journal impact factor (JIF). (For a detailed history of the journal impact factor see *Archambault* and *Larivière* [1.3].)

Garfield did not leave the community without a roadmap. In two short papers introducing the first edition of the JCR—entitled *I. Journals, References and Citations* [1.4], and *II. Why the Journal Citation Reports* [1.5]—*Garfield* provides words of both caution and optimism. Replying to some of the criticism leveled at the Science Citation Index from the scientific community, he provided a justification for interpreting citations as indicators of the *usage* of scholarly literature:

> The more frequently a journal's articles are cited, the more the world's scientific community implies that it finds the journal to be a carrier of useful information. [1.4, p. 1]

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_1

Understanding usage, wrote *Garfield*, would provide critical information on the economics of scholarly publishing and help librarians "counteract the inertia that too often prevails with regard to journal selection" [1.4, p. 1]. Data contained in the JCR would, Garfield argued, provide objective indicators for the use of journals so that librarians could make timely and informed decisions on collection management. The report would provide at scale what had required painstakingly manual analyses in previous decades [1.6]. For researchers, Garfield imagined that the JCR would help them to identify potential venues for publication. Garfield did not advocate for using the JCR to identify elite journals. Rather, he suggested that researchers use the journal-to-journal matrix to identify multidisciplinary venues at "the borders of their own fields." *Garfield* writes [1.5, p. 4–5]:

> the JCR© can be very helpful in deciding where to publish to reach the audience you want to reach. If, for example, you have a paper that deals with some interesting mathematical aspects of biological problems but is nevertheless definitely a biological paper, the JCR© show you which biological journals have the best 'connections' with math, and which are most likely to welcome the paper.

Furthermore, Garfield saw in these new reports the potential to uncover many important dimensions about the nature of science itself. In the conclusion of the introduction to the JCR, *Garfield* states [1.5, p. 5]:

> The use of the JCR can be of far-ranging significance in a field about which I can say least here – science – its planning, its evaluation, its sociology, its history. Citation analysis can be used to identify and map research fronts; to define disciplines and emerging specialties through journal relationships; to determine the interdisciplinary or multidisciplinary character and impact of research programs and projects. I say least about this, to me the most exciting aspect of its potential, because the JCR in its present form is, for such advanced applications, only a sketch of that potential, providing little more than suggestions for further and deeper examination of the massive data bank from which its sections have been extracted.

Garfield concludes with a statement of his hopes: that the JCR will "provide material for innovative research," prompting "imaginative analyses," and stimulate "with every answer it gives more questions that need answers" [1.5, p. 5]. Along these lines, *Garfield* writes in the preface of the first JCR:

> In the introduction I have tried to explain clearly what the JCR is, how it was compiled, how it can be used for some simple purposes for which, I think, it is certainly needed. I have tried also to suggest its usefulness in what I'll call more advanced research. If I have failed in the latter, it is because I have deliberately, and with some difficulty, restrained my own enthusiasm about the value of what some may find at first sight to be merely another handbook of data. Let me say only that the sociology of science is a relatively new field. I believe that JCR will prove uniquely useful in exploring it. [1.7, p. I]

The JCR did indeed provoke a reaction within the research community. Spurred by *Derek de Solla Price*'s call for a *science of science* [1.8], scholars turned to the ISI for data. The JCR and associated products became the backbone for the burgeoning field of scientometrics which sought to address, quantitatively, the questions of science: "its planning, its evaluation, its sociology, its history." In addition to fueling science studies, the JCR found new application alongside the growing emphasis on research evaluation as scholars, institutions, policy-makers, and publishers sought to find ways to measure the success of the research enterprise. This, in turn, had sizeable effects on the science system and scholarly publishing, orienting scholars' research topics and dissemination practices, as well as universities' hiring practices [1.9, 10].

The primary indicator of the JCR—the JIF—has received global attention. As of August 2017, the Core Collection of the Web of Science contained more than 5800 articles that mention the JIF. These papers are not solely in the domain of information or computing science; rather, the majority of papers dealing with JIF are published in scientific and medical journals, demonstrating the pervasive interest in this indicator across scientific fields. The goal of the present chapter is not to summarize this literature per se, but rather to focus on the central limitations that have been raised in the literature and among members of the scientific community.

Drawing on the existing literature as well as on original data, this chapter provides an overview of the JIF and of its uses, as well as a detailed, empirically based, discussion of common critiques. These include technical critiques—such as the asymmetry between the numerator and the denominator, the inclusion of journal self-citations, the length of the citation window, and the skewness of citation distributions—and interpretative critiques—such as the field- and time-dependency of the indicator. Adverse effects of the JIF are discussed and the chapter concludes with an outlook on the future of journal-based measures of scientific impact.

1.2 Calculation and Reproduction

The calculation of the JIF is relatively straightforward: the ratio between the number of citations received in a given year by documents published in a journal during the two previous years, divided by the number of items published in that journal over the two previous years. More specifically, the JIF of a given journal for the year 2016 will be obtained by the following calculation:

$$\frac{\begin{array}{c}\text{Number of citations received in 2016}\\ \text{by items published in the journal}\\ \text{in 2014–2015}\end{array}}{\begin{array}{c}\text{Number of citable items}\\ \text{published in the journal in 2014–2015}\end{array}}$$

Citable items are restricted, by document type, to articles and reviews in the denominator, but not in the numerator [1.11]; an issue we will discuss more in-depth later in the chapter. Therefore, the JIF is generally interpreted as the mean number of citations received by papers published in a given journal in the short term, despite not being exactly calculated as such.

Given its calculation, which uses one year of citation and two years of publication, it combines citations to papers that have had nearly three years of potential citations (i.e., papers published in early 2014) with citations to papers which have had slightly more than a year to receive citations (i.e., papers published at the end of 2015). The JIF is presented with three decimals to avoid ties. However, this has been argued as "false precision" [1.12] with critics advocating for the use of only one decimal point.

Each journal indexed by Clarivate Analytics in the Science Citation Index Expanded (SCIE) and the Social Science Citation Index (SSCI) receives an annual JIF. Given the long half-life of citations (and references) of journals indexed in the Arts and Humanities Citation Index (AHCI), these journals are not provided with a JIF (although some social history journals indexed in the SSCI are included). There has been a steady increase in the number of journals for which JIFs are compiled, in parallel with the increase in indexation. In 1997, 6388 journals had JIFs. This number nearly doubled 20 years later: in 2016, 11 430 received a JIF.

Despite the apparent simplicity of the calculation, JIFs are largely considered nonreproducible [1.13, 14]. However, in order to better understand the calculation of the JIF, we have attempted to recompile, using our licensed version of the Web of Science Core Collection (which includes the Science Citation Index Expanded, Social Science Citation Index, and Arts and Humanities Citation Index), the 2016 JIFs for four journals from the field of biochemistry and molecular biology: Cell, Nature Chemical Biology, PLOS Biology, and the FASEB Journal (of the Federation of American Societies for Experimental Biology). These journals were chosen to represent a range of publishers and open access models while maintaining relative homogeneity in terms of discipline and reputation.

We begin with a careful cleaning of journal names to identify citations that are not automatically matched in the Web of Science (WoS)—that is, citations that bear the name of the journal, but contain a mistake in the author name, volume, or number of pages. The inclusion of these unmatched citations provides the opportunity to essentially reverse-engineer the JIFs presented in the JCR. This reduces the opacity of the JCR, which many consider to be the results of calculations performed on a "separate database" [1.14].

Our empirical analysis (Table 1.1) shows that the inclusion of unmatched citations and the variants under which journal names appear (WoS-derived JIF) provides results that are very similar to the official JCR JIF. This suggests that there is no separate database and one can closely approximate the JIF using only the three standard citation indexes contained in the Core Collection. Furthermore, our results suggest that papers indexed in Clarivate's other indexes—e.g., the Conference Proceedings Citation Index and Book Citation Index—are not included. The inclusion of these databases would lead to an increase of the JIF for most journals, particularly those in disciplines that publish a lower proportion of their work in journals. Most importantly, our analysis demonstrates that with access to the data and careful cleaning, the JIF can be reproduced.

Table 1.1 Citations received, number of citable items, WoS-derived JIF, JCR JIF and proportion of papers obtaining the JIF value, for four journals from the field of biochemistry and molecular biology, 2014–2015 papers and 2016 citations

Journal	Citations		All citations	*N* citable items	WoS-derived JIF	JCR JIF
	Matched items	Unmatched items				
Cell	24 554	2016	26 570	869	30.575	30.410
Nat. Chem. Biol.	3858	356	4214	268	15.724	15.066
PLOS Biology	3331	290	3621	384	9.430	9.797
FASEB Journal	4088	802	4890	881	5.551	5.498

1.3 Critiques

The JIF has been called a "pox upon the land" [1.9], "a cancer that can no longer be ignored" [1.15], and the "number that's devouring science" [1.9]. Many scholars note the technical imperfections of the indicator—skewness, false precision, absence of confidence intervals, and the asymmetry in the calculation. Considerable focus has also been paid to the misapplication of the indicator—most specifically the use of the indicator at the level of an individual paper or author [1.16]. We will not review this vast literature here, much of which appears as anecdotes in editorial and comment pieces. Instead, we provide original data to examine the most discussed technical and interpretive critiques of the JIF. Furthermore, we provide new information on a previously understudied dimension of the JIF—that is, the inflation of JIFs over time.

1.3.1 The Numerator/Denominator Asymmetry

Scholarly journals publish several document types. In addition to research articles, which represent the bulk of the scientific literature, scholarly journals also publish review articles, which synthesize previous findings. These two document types, which are generally peer-reviewed, account for the majority of citations received by journals and constitute what Clarivate labels *citable items*. Over the 1900–2016 period, 69.7% of documents in the Web of Science were considered as citable items. This proportion is even more striking for recent years, with 76.0% of documents published in 2016 labeled as citable items. Other documents published by scholarly journals, such as editorials, letters to the editor, news items, and obituaries (often labeled *front material*), receive fewer citations, and are thus considered *noncitable items*. There is, however, an asymmetry in how these document types are incorporated into the calculation of the JIF: while citations received by all document types—citable and noncitable—are counted in the numerator, only citable items are counted in the denominator. This counting mechanism is not an intentional asymmetry, but rather an artifact of method for obtaining citation counts. As mentioned above, to account for mistakes in cited references and to try to be as comprehensive as possible, Clarivate focuses retrieval on all citations with the journal name or common variant [1.17] rather than using a paper-based approach to calculating citations. This has the effect of inflating the JIF: citations are counted for documents which are not considered in the denominator. The variations in document types (i.e., reduction of the number of citable items in the denominator) has also been argued as the main reason for JIF increases [1.18].

To better understand the effects of document types on the calculation of the JIF, we compiled, for the sample of four journals from the field of biochemistry and molecular biology, as well as for *Science* and *Nature*—both of which publish a high percentage of front material—citations received by citable items, noncitable items, as well as unmatched citations (Table 1.2). Following *Moed* and *van Leeuwen* [1.19, 20], our results show that noncitable items and unmatched citations account for a sizeable proportion of total citations received, from 9.8% in the case of *Cell* to 20.6% in the case of *FASEB Journal*. For the four journals from biochemistry and molecular biology, unmatched citations account for a larger proportion of citations than noncitable items. Given that these unmatched citations are likely to be made to citable items, this suggests that, at least in the case of disciplinary journals which do not typically have a large proportion of front material, the asymmetry between the numerator and the denominator does not inflate JIFs in a sizeable manner. The effect of noncitable items is much greater for interdisciplinary journals such as *Science* and *Nature*. As shown in Table 1.2, for both *Nature* and *Science*, more than 5000 citations are received in 2016 by noncitable items published in the journal in 2014–2015. This accounts

Table 1.2 Number and proportion of citations received by articles, reviews, noncitable items, and unmatched citations, for four journals from the field of biochemistry and molecular biology, as well as *Nature* and *Science*, 2014–2015 papers and 2016 citations

Journal	Articles		Reviews		Noncitable items		Unmatched citations		*N* citable items	Symmetric impact factor	JCR impact factor	Increase (%)
	N	(%)	*N*	(%)	*N*	(%)	*N*	(%)				
Cell	20 885	78.6	3068	11.5	601	2.3	2016	7.6	869	27.564	30.410	10.3
Nat. Chem. Biol.	3263	77.4	378	9.0	217	5.1	356	8.4	268	13.586	15.066	10.9
PLOS Biology	3088	85.3	6	0.2	237	6.5	290	8.0	384	8.057	9.797	21.6
FASEB Journal	3650	74.6	235	4.8	203	4.2	802	16.4	881	4.410	5.498	24.7
Nature	55 380	78.6	3925	5.6	5067	7.2	6047	8.6	1784	33.243	40.140	20.7
Science	45 708	73.0	4886	7.8	5657	9.0	6340	10.1	1721	29.398	37.210	26.6

for 7.2% and 9.0% of citations, respectively, which is greater than the percentages obtained by the sample of disciplinary journals (2.3–6.5%). Results also show that the difference in the "symmetric" JIF—with only citable items in the numerator and denominator—and the JCR JIF is greater for *Nature* and *Science* than *Cell* or *Nat. Chem. Biol.*, mostly because of citations to nonsource items. However, at scale—i.e., all journals having a JIF in 2016—the relationship between the JIF and the symmetric impact factor is quite strong, with an R^2 of 0.96 (Fig. 1.1).

These results demonstrate that the asymmetry has different effects based on (1) the proportion of front material, and (2) the completeness of citations received by the journal. Moreover, they show that most of the additional citations—i.e., citations not directly linked to citable items—are unmatched citations rather than direct citations to noncitable items. Given most of these unmatched citations are likely to be directed at source items, a more accurate calculation of the JIF could exclude citations to nonsource items, but retain unmatched citations. Of course, the ideal solution would be to perform additional data cleaning to reduce the proportion of unmatched citations and have perfect symmetry between the numerator and denominator.

1.3.2 Journal Self-Citations

The inclusion of journal self-citations in the calculation of the JIF has been a cause for concern, as it opens the door for editorial manipulations of citations [1.21–23]. Journal self-citations are those citations received by the journal that were made by other papers within that same journal. This should not be conflated with self-references, which is the proportion of references made in the articles to that journal. This is a subtle, but important difference: the proportion of self-citations is an indication of the relative impact of the work on the broader community, whereas the proportion of self-references provides an indication of the foundation of work upon which that journal is built. From a technical standpoint, the main concern in the construction of the JIF is the degree to which self-citations can be used to inflate the indicator. Given that self-citations are directly under the control of the authors (and, indirectly, the editors), this has been seen as a potential flaw that can be exploited by malicious authors and editors.

There are many myths and misunderstandings in this area. For example, it has been argued that authors in high-impact journals are more likely to self-cite than those in low-impact journals because "the former authors in general are more experienced and more successful" [1.13, p. 50]. However, this is a conflation of self-citations and self-references. Authors with longer publication histories are, indeed, more likely to have material to self-reference. However, successful authors are likely to have lower self-citation rates, as they are likely to generate citations from a broader audience. Furthermore, this conflates the practices of an individual author (who publishes in many journals) to the self-citation of a journal, which is much more dependent upon the specialization of the journal, among other factors [1.24]. There is also a distinction to be made between the number and proportion of self-citations. As ISI observed in internal analyses, "a high number of self-citations does not always result in a high rate of self-citation" [1.25, par. 15]. For example, a study of psychology journals found that articles in high-impact journals tend to receive a higher *number* of self-citations than articles in lower impact journals. However, the *ratio* of self-citations to total citations tends to be lower for high-impact journals [1.13].

Producers of the JIF thus face a Cornelian dilemma when it comes to self-citations: while including them can lead to manipulation, excluding them penalizes niche journals and certain specialties. In response to these concerns, ISI undertook an analysis of the prevalence and effect of journal self-citations [1.25]. In an analysis of 5876 journals in the 2002 Science Edition of the JCR, ISI found that the mean self-citation rate was around 12%. Our analysis of 2016 citation data for papers published in 2014–2015 reinforces this: we find that the percentage of self-citations across all disciplines remains around 12% (Fig. 1.2). However, the percentage varies widely by discipline, with arts and humanities having far higher degrees of self-citation than clinical and biomedical research. This suggests that, on average, the majority of citations do not come in the form of self-citations and makes abuses easier to identify.

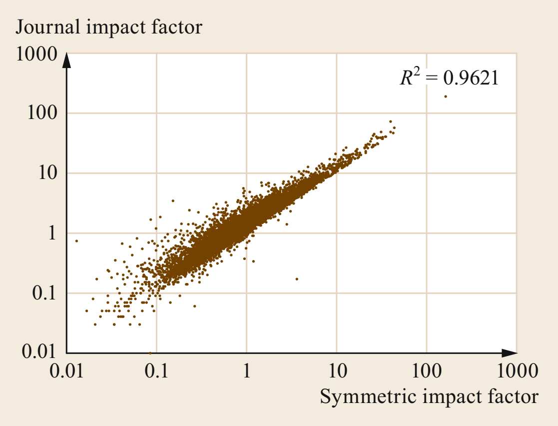


Fig. 1.1 Correlation between the JIF and the symmetric impact factor, 2016

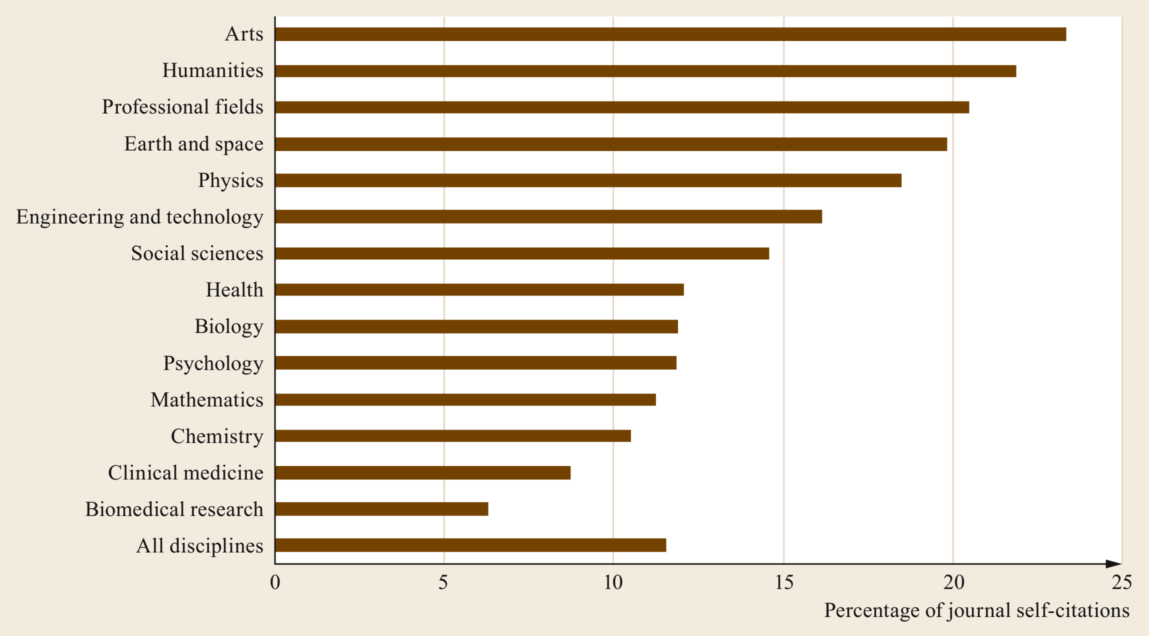


Fig. 1.2 Percentage of journal self-citations, by discipline, for citations received in 2016 by papers published in 2014–2015

The ISI analysis also examined the correlation between self-citation rates and JIFs. While studies focusing on particular domains have found varying results [1.13, 26, 27], the large-scale analysis by ISI found a weak negative correlation between JIF and rates of journal self-citation [1.25]. The analysis noted that self-citation had little effect on the relative ranking of high-impact journals, given that journals in the top quartile of JIFs tended to have self-citation rates of 10% or less. Lower impact journals, however, were more dependent upon self-citations [1.25]. We found similar results for all 2016 journals. As shown in Fig. 1.3a, there is a relatively strong correlation between a journal's total number of external citations (i.e., non self-citations) and its number of self-citations, which suggests that external- and self-citations are related, but also that there are other factors influencing the relationship, such as the level of specialism of the journal. For instance,

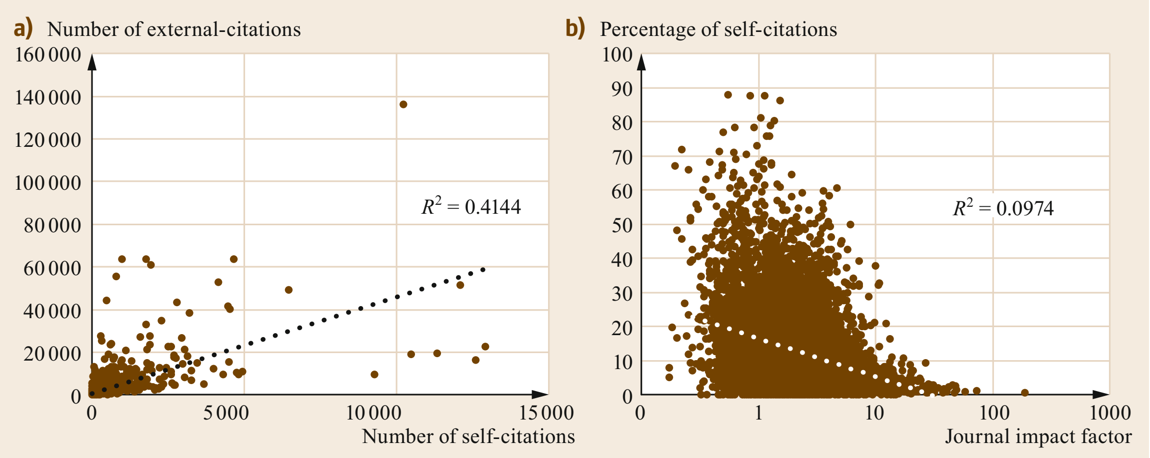


Fig. 1.3a,b Correlation at the journal level between **(a)** the number of journal *external* citations and number of journal self-citations and **(b)** the percentage of self-citations and the Impact Factor, for year 2016. Only journals with at least 50 citations in 2016 to material published in 2014–2015 are shown.

2014–2015 papers from the *Journal of High Energy Physics* received 18 651 citations in 2016, of which 9285 (50%) came from the same journal. Other more generalist journals in that domain—such as *Physical Review B* and *Monthly Notices of the Royal Astronomical Society*—exhibit a similar pattern.

The irony of the concern between self-citation and JIFs, however, is that the relationship is inverted: there is actually a negative relationship between the percentage of self-citations for a journal and the JIF (Fig. 1.3b). That is, those journals with the highest JIFs tend to have the lowest percentage of self-citations. There is, simply speaking, a limit on the advantages of self-citations. There are many more articles outside of the journal than within and relying on citations within can only generate a finite number of citations. A variant JIF omitting self-citations is now available in the JCR. However, the two-year JIF including self-citations continues to be the dominant form.

1.3.3 Length of Citation Window

The JIF includes citations received in a single year by papers published in the journal over the two preceding years. As such, it is generally considered to cover citations received by papers over a two-year window. This focus on the short-term impact of scholarly documents is problematic as it favors disciplines that accumulate citations faster. For example, comparing mean citation rates of papers published in the *Lancet* and in the *American Sociological Review* (ASR)—two journals with very different JIFs (47.83 versus 4.4 in 2016)—*Glänzel* and *Moed* [1.29] have shown that while papers published in the *Lancet* had a higher mean citation rate for two- and three-year citation windows, those published in *ASR* were more highly cited when a longer citation window was used.

This trend can be observed at the macrolevel: Figure 1.4 presents the annual number of citations (Fig. 1.4a), cumulative number of citations (Fig. 1.4b), and the cumulative proportion of citations (Fig. 1.4c), for all papers published in 1985 across four disciplines (biomedical research, psychology, physics, and social sciences). These data show that citations to biomedical research and physics peak two years following publication, while citations are relatively more stable following publication year in psychology and the social sciences. It is particularly revealing that psychology papers receive, on average, more citations (cumulatively) than physics papers. While physics papers generate more citations than psychology papers within the first five years, the reverse is true for the following 25 years.

Despite these disciplinary differences in the speed at which citations accumulate, the two-year window appears to be ill-suited across all disciplines, as it covers only a small fraction of citations received over time. For example, using a 30-year citation window, we find that the first two years captures only 16% of citations for physics papers, 15% for biomedical research, 8% for social science papers, and 7% in psychology. Figure 1.4

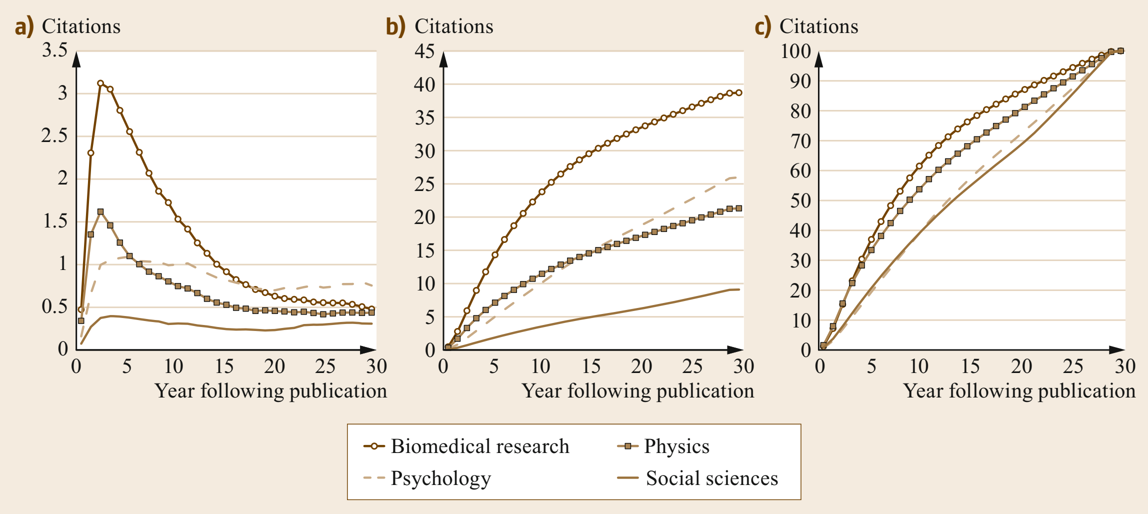


Fig. 1.4 Number of citations (**a**), cumulative number of citations (**b**), and cumulative proportion of citations (**c**), by year following publication for papers published in 1985 in biomedical research, psychology, physics and social sciences (according to the National Science Foundation (NSF) field and subfield classification (after [1.28])

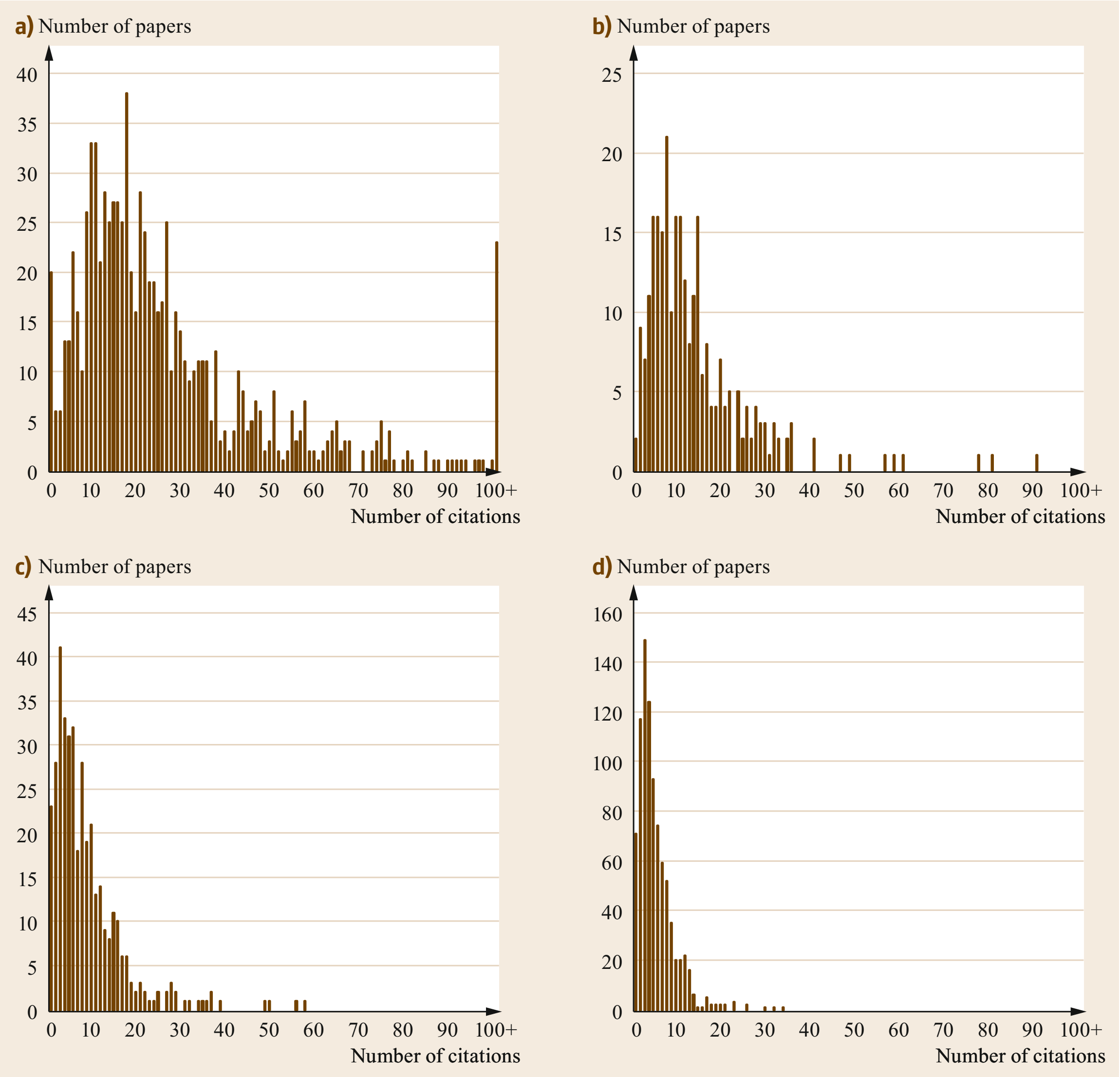


Fig. 1.5a–d Distribution of citations received by articles and reviews, for four journals from the field of biochemistry and molecular biology, 2014–2015 papers and 2016 citations. **(a)** Cell (JIF = 30.410), **(b)** Nat. Chem. Biol. (JIF = 15.066), **(c)** PLOS Biology (JIF = 9.797), **(d)** FASEB Journal (JIF = 5.498)

also shows that papers in biomedical research accumulate citations faster than in the other three domains. For instance, they accumulate 50% of their citations in the first eight years following publication, while it takes nine years for physics papers, 13 years for psychology papers, and 14 years for social science papers to reach the same threshold. In order to take such differences into account, the JCR has provided, since 2007, a 5-year JIF. Despite this improved citation window, which provides a more complete measurement of the impact of papers and journals, the two-year JIF remains the gold standard.

1.3.4 Skewness of Citation Distributions

Nearly a century of research has demonstrated that science is highly skewed [1.30] and that productivity and citedness are not equally distributed among scholars, articles, institutions, or nations. It is perhaps of little surprise, therefore, that the citedness of articles within a journal is also highly skewed. This was the main premise of an article published in 1992 by *Per O. Seglen* [1.31], who produced a robust empirical analysis demonstrating that a minority of papers in a journal accounted for the vast majority of citations. Given this

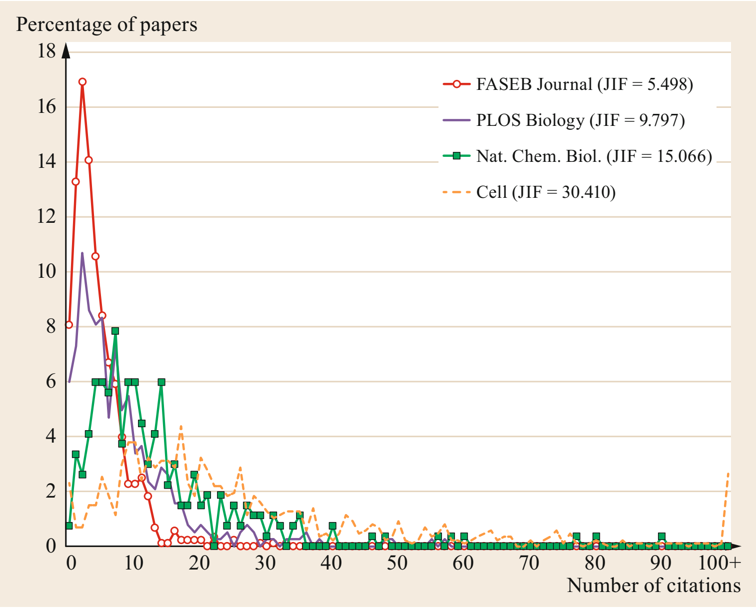


Fig. 1.6 Relative distribution of citations received by articles and reviews, for four journals from the field of biochemistry and molecular biology, 2014–2015 papers and 2016 citations

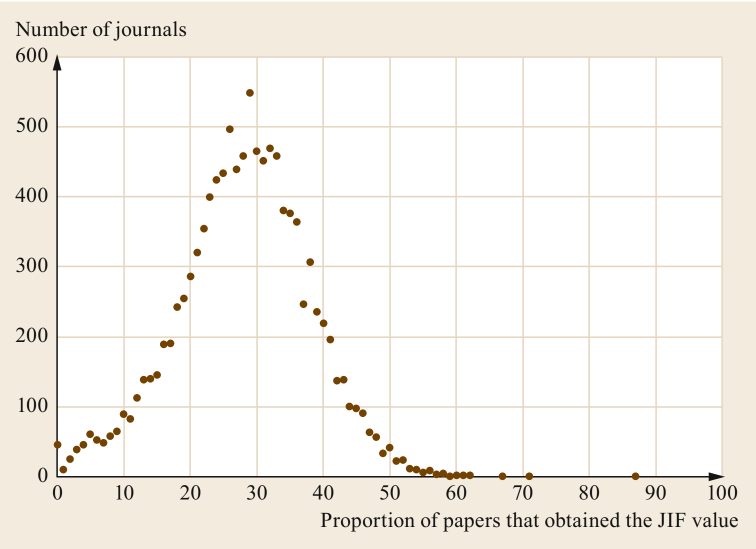


Fig. 1.7 Distribution of the number of journals, by proportion of papers that obtained the JIF value, 2014–2015 papers and 2016 citations

skewness in the citation distribution, Seglen argued that the JIF was unsuitable for research evaluation.

To illustrate this skewness, we provide—for the four biochemistry and molecular biology journals mentioned above—the distribution of citations received in 2016 by papers published in 2014–2015, both as absolute values (Fig. 1.5) and as percentages of papers (Fig. 1.6). It shows that, for all journals, most of the papers have a low number of citations and only a few obtain a high number of citations. Of course, the distribution for *Cell*—with a JIF of 30.410—is more right-skewed than *FASEB Journal*—which has a JIF of 5.498—but despite this, their citation distributions still have sizeable overlap, as shown in Fig. 1.6. Also striking is the similarity of the skewness: for all of these four journals, a nearly identical percentage of papers—28.2–28.7%—obtain a citation rate that is equal or greater to the JIF for that journal.

Extending the analysis across all journals indexed in the 2016 JCR confirms this pattern (Fig. 1.7). There is a fairly normal distribution when plotting journals by the percentage of their papers that obtain the corresponding JIF value or above. As shown, the vast majority are around 30%. Nearly 73% of the journals fall between 20–40%. Only in 1.3% of journals ($N = 141$) do at least 50% of the articles reach the JIF value. This fundamental flaw in the calculation—to compile an average on a nonparametric distribution—has been heavily discussed in the literature [1.32] as both a statistical aberration and also for the common misinterpretation: to use the JIF as an indicator at the article or individual level. Our analysis demonstrates the fairly weak predictive power of the JIF—that is, one cannot extrapolate from the impact factor of the journal to the potential citedness of the article as only one-third of the articles are likely to obtain that value. There have been many suggestions to account for the skewness, such as compiling a median-based JIF [1.33, 34] or reporting citation distributions [1.32]. However, contrary to other alternatives (such as the 5-year JIF and JIF excluding self-citations), no alternatives have been adopted by the JCR to address this limitation.

This is not to say, of course, that there is no relationship between JIF and future citedness. For example, using identical papers published in journals with different JIFs, *Larivière* and *Gingras* [1.35] found that the mean number of citations of the paper published in the journal with the highest JIF obtained twice as many citations as its twin published in the journal with the lowest JIF. However, the relationship between the JIF and the citedness of the articles has weakened over time: as shown by *Lozano*, *Larivière* and *Gingras* [1.36] using Web of Science data—and confirmed by *Acharya* [1.37] using Google Scholar—the correlation between the JIF and article-level citations has been decreasing since the mid-1990s. One potential explanation for this is the changing referencing practices of scholars. Citations are less concentrated over time [1.38] and scholars are citing increasingly older literature [1.39] and, as they do, more of the citations fall out of the two-year citation window of the JIF.

It would be irresponsible here not to mention the *Lucas* critique [1.40], which argues against predicting the effects of policy changes based on aggregated historical data. The Lucas critique was developed for economic data, but has wide applicability for the social sciences. In bibliometrics, one should be wary of making predictions about future citations, based on the past performance of scholarly objects. Referencing and citing patterns vary over time as do the sociopolitical factors of scholarship. Furthermore, the construction of citation indicators changes behavior (as we discuss later in this chapter). Therefore, we caution against making predictions with citation data.

1.3.5 Disciplinary Comparison

Field differences in citations are well established and field-normalized indicators have been the norm for several decades [1.41, 42]. However, the JIF is not among these. The simplicity of the calculation fails to normalize for the vast differences in citing practices across disciplines, such as the number of references per document and age of references. As shown in Table 1.3, disciplines that publish papers with longer cited reference lists—especially in terms of WoS-indexed papers—generally have higher JIFs than those with shorter lists. Furthermore, disciplines that cite more recent material—which fall in the JIF two-year citation window—are more likely to have higher JIFs than those that cite older material.

These differences also highlight the importance of references to other WoS-indexed material (source items), which are those that are taken into account in the compilation of the JIF. For instance, while the mean number of references in biology and biomedical research are almost identical, the mean JIF of journals in biology is less than half of those in biomedical research.

Table 1.3 Mean and maximum JIF of journals, mean number of cited references per paper (all material and only to WoS source items), and mean age of cited literature, by discipline, 2014–2015 papers and 2016 citations

Discipline	Mean JCR JIF	Maximum JCR JIF	Mean N references	Mean N references to WoS source items	Mean age of cited literature
Biology	1.683	22.81	48.99	34.45	14.72
Biomedical research	3.526	46.60	48.94	43.19	10.26
Chemistry	2.768	47.93	46.37	41.31	10.37
Clinical medicine	2.976	187.04	41.94	34.78	9.77
Earth and space	2.173	30.73	53.71	38.67	13.06
Engineering and technology	1.989	39.74	36.35	24.77	10.44
Health	1.647	17.69	39.08	24.52	9.86
Mathematics	1.017	9.44	26.56	16.53	16.65
Physics	2.699	37.85	36.57	29.58	12.55
Professional fields	1.565	11.12	53.51	27.68	13.09
Psychology	2.050	19.95	54.56	38.30	13.00
Social sciences	1.199	6.66	49.09	21.74	15.12

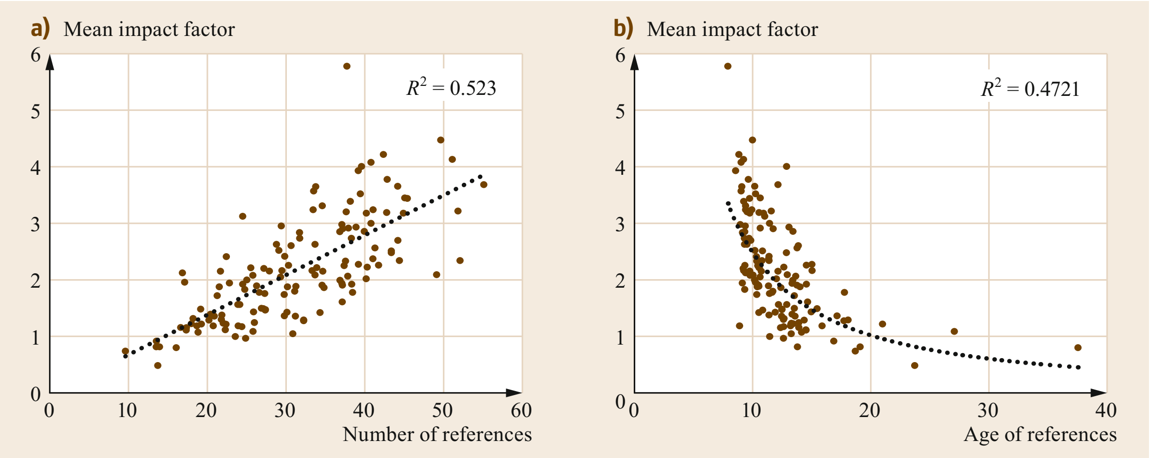


Fig. 1.8a,b Correlation between the journal impact factor and number of cited references to WoS source items **(a)** and age of references **(b)**, by NSF speciality, 2014–2015 papers and 2016 citations

This difference is explained by the fact that a large proportion of references made by biology journals do not count in the calculation of JIF as they are made to non-WoS (and, thus, JCR) material, while the vast majority of references of biomedical research journals are to WoS-indexed journals.

The same patterns are observed at the level of NSF specialities (Fig. 1.8). Specialties that cite a higher number of references per paper on average typically have higher JIFs (Fig. 1.8a), as are specialities that cite younger material (Fig. 1.8b). Therefore, the indicator cannot be used to compare across disciplines: medical researchers are much more likely to publish in journals with high JIFs than mathematicians or social scientists, and this is strictly due to different disciplines' publication and referencing practices rather than anything that relates to the scholarly impact of the journal.

1.3.6 Journal Impact Factor Inflation

While the calculation of the JIF has remained stable, values obtained by journals have not. The average JIF value has increased over time, both as a function of the number of papers in existence and the increasing length of their reference lists [1.39]. In 1975, the journal with the highest JIF was the *Journal of Experimental Medicine*, with a JIF of 11.874. In the 2016 JCR edition, the highest JIF was 187.040 for *CA: A Cancer Journal for Clinicians*. As shown by Fig. 1.9, a general inflation of the JIF has been observed over the last 20 years. For instance, while only 49 journals (0.8% of total) had a JIF above 10 in 1997, this increased to 105 (1.3%) in 2007, and to 201 (1.8%) in 2016. Average JIF values have increased from 1.125 in 1997, to 1.707 in 2007, and then to 2.178 in 2016. Of course, not all journals have observed these increases. One notable example is *PNAS*, which has remained quite stable—the 1975 JIF was 8.989 and, despite some intermittent increases, was only slightly higher at 9.661 in 2016.

The inflation of the JIF across time is an important element for interpretation. Many editors wait with baited breath for the release of the next JIF: increases are celebrated as an accomplishment of the editor and the journal [1.43–45]. Moreover, publishers, such as *Elsevier* [1.46], *Springer* [1.47], and *Wiley* [1.48] publicize their JIF increases with little to no conversation about the expected inflation rates. For example, the Wiley press release boasts that 58% of Wiley journals increased their JIFs between 2014 and 2015. What the press release fails to note is that 56% of all journals in the JCR increased during that same time period. Of course, reporting a relative increase is much less persuasive. As there is no established mechanism for acknowledging inflation in reporting, editors and publishers continue to valorize marginal increases in JIFs which have little relation to the performance of the journal.

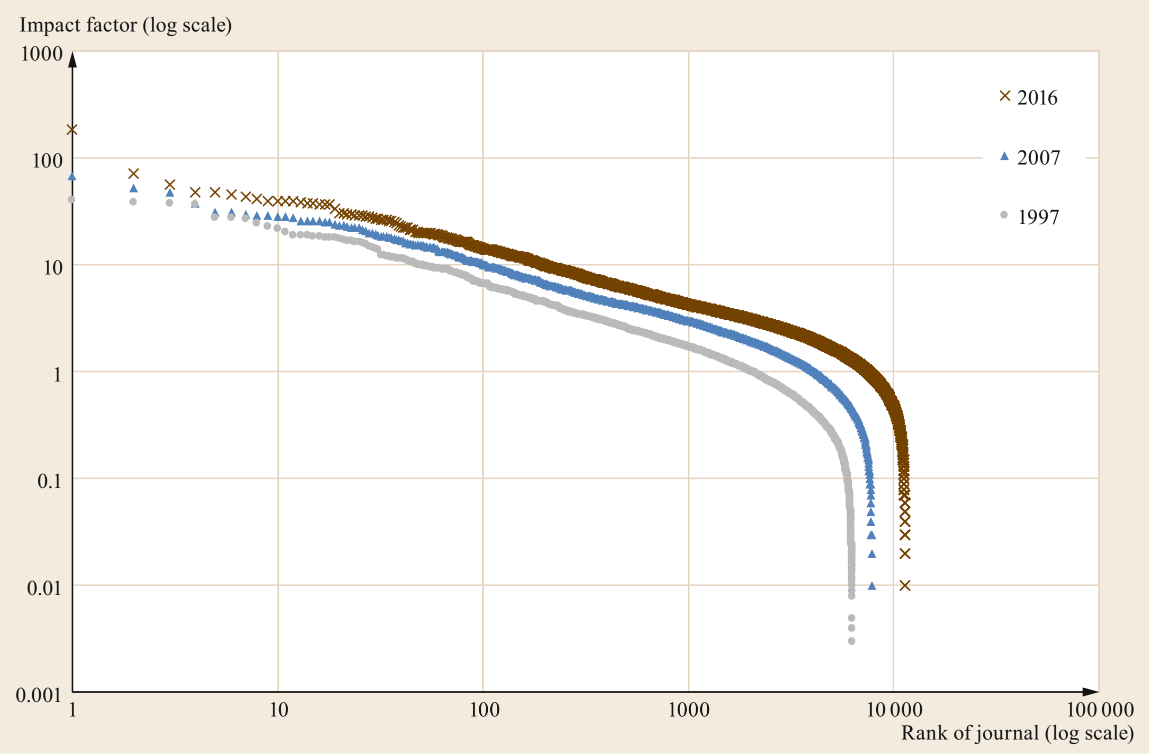


Fig. 1.9 Impact factor by journal as a function of rank, for years 1997, 2007, and 2016

1.4 Systemic Effects

There is no doubt that a political economy has emerged around citation indicators. Nearly two decades ago, *Sosteric* [1.49, p. 13] commented on

> the neoliberal need for surveillance, the push for administrative measures of scholarly performance and productivity, [and] the growing need for post-publication measures of scholarly impact.

He did not characterize scholars as resisters of this panopticon, but rather as adaptive actors in the system. Adaptation for survival and success is well-known across all fields of science: research evaluation is no different. Several scholars have warned against the negative consequences of constructing indicators of social activities [1.50–52]. As *Cronin* and *Sugimoto* summarized [1.53, p. 751]:

> The use of metrics, whether to monitor, compare or reward scholarly performance, is not a value-neutral activity. Metrics are shaped by, and in turn shape policy decisions; they focus the institutional mind, influence the allocation of resources, promote stratification and competition within science, encourage short-termism and, ultimately, affect the ethos of the academy... As reliance on metrics grows, scholars, more or less consciously, alter the way they go about their business; that is, their behaviors, motivations and values change, incrementally and unwittingly perhaps, as they adapt to the demands and perceived expectations of the prevailing system.

While it would be beyond the scope of the chapter to detail all the systemic effects of scholarly indicators, we focus on the negative and often intentionally malicious effects related to the use and promotion of the JIF. Specifically, we discuss JIF engineering, its relationship with institutional evaluation policies, the application of JIF for evaluating individual researchers and papers, and the creation of imitation indicators.

1.4.1 Journal Impact Factor Engineering

In a context where the JIF determines the fate of a journal—from submission rates to pricing—some editors and publishers have developed subterfuges to increase their JIF which, in turn, decreases the validity of the indicator. Such stratagems aimed at *artificially* increasing impact factors have been called *journal impact factor engineering* [1.22]. One well-documented tactic is to prey on the asymmetry in the calculation and to publish more *front material*—such as editorials, letters to the editor, etc., which are considered by Clarivate as noncitable items [1.22]. Another similar approach is to cite the home journal excessively in editorials and other front matter [1.22]. For example, many journals publish annual *highlights* or other documents with a high number of internal references [1.27]. Whether malicious or not, these documents unduly inflate—and thereby invalidate—the JIF.

A more subversive approach has been to engage in citation coercion or cartels [1.9, 23, 54–56]. The expression *citation cartel* is largely attributed to *Franck* [1.57], who used it to refer to the ways in which monopoly power is exercised by publishers and editors on authors in scientific publishing, and noted the complicity of authors who act as *citation-maximizers* in the scholarly communication system. This complicit behavior has been empirically demonstrated: in a study of nearly 7000 scholars, the majority reported that they would acquiesce to editorial coercion in order to get published [1.58]. The same study also showed that 20% of these scholars said they had been subject to coercive self-citation—that is, requests from editors to add references to irrelevant papers published within the journal [1.58]. An expansion of this study—with new disciplines added—placed this rate at 14.1%. Both the initial and follow-up studied confirmed that coercion was more common among higher impact journals [1.58, 59].

Faced with accusations of extortion [1.9], editors will often argue the innocence of and scientific rationale for these citations [1.60]. However, several editors themselves have been caught engaging in JIF boosting, by excessively citing their own journal in editorials [1.22]. There are also egregious examples of coercion. For example, in 2017, the editor of the journal *Land Degradation & Development*—who sat on the board and reviewed for other journals in the field—took advantage of his positions to increase the JIF of his own journal. Among the 82 manuscripts he handled as an editor and reviewer for other journals, he suggested 622 additional references, almost exclusively to the journal of which he was Editor-in-Chief [1.61]. The result was an astronomic rise in the JIF of the journal he edited, from 3.089 to 8.145 between 2014 and 2015. These flagrant abuses signal that editors are highly aware of the benefits derived from these manipulations.

Coercive self-citation is easier to identify than citation-stacking, which has become synonymous with the contemporary notion of *citation cartels*. There can be several legitimate explanations for tightly coupled exchange of citations between journals, particularly in highly specialized fields. However, when these exchanges are done with the explicit intent of increasing the citedness of the journal, these are referred to as citation cartels. Although there have been a few attempts to identify cartels [1.62–64], detection is difficult on a number of fronts. Technically, the ability to identify cartels becomes more difficult as the size of the cartel increases. Furthermore, the notion of a cartel implies intentionality and premeditation—something that is impossible to prove using bibliometric data alone.

Thomson Reuters (and, subsequently, *Clarivate Analytics*) has worked to police inappropriate citation activity—though they note that they do not "assume motive on behalf of any party" [1.65]. Each year, Clarivate provides a report of titles suppressed due to "anomalous citations patterns" and the reason for removal from the JCR [1.66]. Journals can be removed due to excessive self-citation or citation stacking, although thresholds are considered to be *extremely high* [1.67]. For example, in an analysis in 2002, the Institute for Scientific Information (the precursor to Thomson Reuters and Clarivate) found that for 82% of their titles, self-citation rates were at or below 20% [1.25]. It is assumed, therefore, that all journals will engage to some degree in self-citation. However, when the proportional increase in the JIF is due largely to an increase in self-citation, the journal is flagged for further analysis [1.65]. This is not an entirely uncommon practice and represents the dominant reason for suppression from the JCR. Perhaps as a result of reporting, cases of citation stacking have decreased over time [1.68].

Other scholars have also sought to create indicators for identifying excessive self-citations: *Chorus* and *Waltman* [1.69] created the Impact Factor Biased Self-citation Practices (IFBSCP) indicator to examine the relationship between the share of self-citations for the years included in the impact factor to those in the preceding five years. To validate this as an indicator of coercive self-citations, they examined the rates of IFBSCP for the 64 journals identified in *Wilhite* and *Fong* [1.58] as engaging in coercive citation behavior. They found that the named journals had IFBSCP rates 25% higher than the average social science journal, which suggests that their indicator measure is related to coercive behavior. This suggests that indicators may be developed to help identify—and hopefully curb—inappropriate citation behavior.

1.4.2 Role of Evaluation Policies

Impact factor engineering does not happen in a void: these actions are a consequence of evaluation policies and practices. Institutions and individuals are complicit actors in promoting the JIF in a research evaluation context. Although soft persuasions towards maximizing impact can be seen across the scientific system, they are made most manifest in the cash-based reward systems, such as those documented and publicized in China. Chinese policies offering financial reward based on WoS-indexed publications began in earnest in the 1990s, to motivate production and increase international visibility [1.70]. However, as noted by other studies [1.71], increasing national production does not necessarily equate to an increase in citedness, and might actually lead to a decrease. Therefore, China has moved steadily away from publication-based incentives in favor of citation-based indicators, particularly those based on JCR-quartiles of JIFs [1.70]. At face value, these policies seem well-intentioned and even laudable—encouraging quality over quantity. However, given that the cash award for a *Nature* or *Science* article can be 20 times an annual salary in China [1.70], these rewards can create strong incentives for inappropriate behavior. Although one cannot determine causality, the rise in fraudulent authorship, data falsification, and data fabrication in China [1.72] in parallel with these rewards is disconcerting. There is even evidence of an industry of authorship for sale in China, in which authorship is sold to scholars at rates that often exceed salaries [1.73].

Furthermore, cash incentive programs have been correlated with increased submission, but not with publication [1.74]. Although most authors are fairly efficient at selecting appropriate journals, many authors tend to submit to higher impact factor journals first and then resubmit down the JIF ladder until they find an acceptance [1.9, 75]. Increasing the pressure to submit to high-impact factor journals creates a burden on the scientific system and slows the pace of science as editors and reviewers are tasked with reviewing papers that are not submitted to the most appropriate venues. On a more fundamental level, financial rewards for papers externalizes the incentive to do scientific work. This contradicts central ideals of scholarship, in which scholars should be free from external pressures [1.76]. A reward more than 20 times an annual salary inverts the reward system—prioritizing external (i.e., economic capital) over intrinsic (academic capital) rewards.

There is also a danger in tying rewards to publication in particular journals. The most appropriate venue for many scholars—particularly those in the social sciences and humanities—may not be in a WoS-indexed publication at all. By emphasizing JIFs, the coverage biases of the WoS become prioritized [1.77]; that is, journal articles in the natural and medical sciences published in English are particularly incentivized. Some have argued that switching to English-language journals increases the visibility of science produced in countries where English is not the dominant language [1.78, 79]. However, others have expressed concern about the effects of a monolingual scholarly publishing industry [1.80]. For instance, *Larivière* [1.81] has shown that Canadian scholars in the social sciences and humanities were three times less likely to publish on Canada-related research topics when publishing in US journals than in Canadian journals, which demonstrates how journal venues directly affect the type of research performed.

1.4.3 Application at the Individual Level

As the JIF is based on a skewed distribution and, thus, is a weak predictor of individual papers' citation rates, its use as an indicator of the *quality* of individual researchers and papers—sometimes labeled the ecological fallacy [1.82]—is perhaps the most egregious misappropriation of the indicator. As Anthony van Raan noted (quoted in [1.83, p. 864–865]):

> if there is one thing every bibliometrician agrees, it is that you should never use the JIF to evaluate research performance for an article or for an individual—that is a mortal sin.

A less hyperbolic, but similarly unequivocal statement can be found from other bibliometricians: *Henk Moed* noted that such measures "have no value in assessing individual scientists" [1.84]. Despite these admonitions, the JIF is increasingly used as an indicator to evaluate individual scholars (among others, [1.70, 85]). While some might argue that publication in a journal with a high JIF is itself an achievement, given the relatively lower acceptance rates of these journals, the concern is more about the equation of the value of an article or individual with the past ranking of a journal [1.86]. This can lead to gross goal displacement [1.87], in which scholars tailor their topics for certain indicators.

Scholars are increasingly "thinking with indicators"—that is, allowing indicators to guide the process of science-making [1.10]. Specifically, scholars choose topics and dissemination venues not on scientific bases, but rather to meet certain incentive structures. In doing so, scholars substitute a "taste for science" with a "taste for rankings" [1.87]. This is not a particularly novel claim. As early as 1991, *Holub* and colleagues noted that "WHERE a scientist published has become much more important than WHAT he is publishing [capitalization in original]" [1.88]. However, the impact factor obsession [1.12] has grown to the level where some scholars would rather destroy a paper than publish below a certain JIF threshold [1.89]. This has led to a complicated and cyclical relationship between JIF, value, and reputation that is increasingly internalized into the process of scholarship [1.10].

Scholars are aware of these negative effects: several initiatives in recent years have sought to disentangle journal rankings from individual rankings. At the 2012 annual meeting of the American Society for Cell Biology (ASCB), a group of editors and publishers produced the san francisco declaration on research assessment, colloquially referred to as DORA [1.90]. The declaration called for the elimination of the use of JIFs for assessment of individual scholars and articles [1.90], stating that the JIF was not appropriate "as a surrogate measure of the quality of individual research articles, to assess an individual scientist's contribution, or in hiring, promotion or funding decisions" [1.90, 2]. As of July of 2017, the declaration had nearly 13 000 individual signers and nearly 900 organizational signers. Funding agencies have also responded: the National Health and Medical Research Council (NHMRC) in Australia produced a statement unequivocally denouncing JIFs for evaluating individual papers [1.91] and discontinued reporting of JIFs for evaluation. Nobel laureates and other high-profile scholars have also spoken out against JIFs [1.92] and boycotted high-impact factor journals [1.93]. However, these are privileged boycotts and resistance is much more difficult for those who are not well-established in the scientific system.

1.4.4 Knock-Off Indicators

The JIF has become a brand and, like any other luxury good, there is an industry of imitation. In recent years, a cottage industry of fake impact factors has emerged, with strong ties to predatory publishers. Librarian Geoffrey Beall—who for many years ran the well-known and controversial list of predatory publishers—identified more than 50 organizations that provide *questionable* or *misleading* metrics at the researcher, article, and journal level [1.94]. The complicated web of mimicry is difficult to disentangle: the names of the organizations often replicate the name or acronym of the Institute for Scientific Information—e.g., the *Institute for Science Information (ISI)*, the *Index Scientific Journals (ISJ)*, or the *International Scientific Indexing (ISI)*—or the JIF—e.g., the *Journal Influence Factor-JIF*, the *General Impact Factor*, or the *Science Impact Factor*. One organization even goes as far as to imitate both the name of the indicator and that of the organization: journals can apply to the "Global Institute for Scientific Information (GISI)" to obtain a "journal impact factor" [1.95]. Several journals seem to have either fallen prey or are complicit in this deceit: for instance, the list of journals to which GISI has attributed a "journal impact factor" increased from 24 in 2010 to a high of 668 in 2011–2013. The numbers have been steadily dwindling, but there are still 153 journals listed in 2016. The listed journals come from both predatory and well-established publishers.

The organizations often go to lengths to maintain their deceit. For example, one website includes a red pop-up box warning editors and publishers that another company is scamming the original predatory company. The text reads:

> This is to inform you that somebody is using our name (International Impact Factor Services) to deposit the fee for Impact Factor &he saying that he show your impact factor in our website, but do not reply those mails. If you answer those mails you will responsible for that [1.96].

This is not the only bait and switch in the impact factor market. For example, one of the only published articles on fake JIFs was published in *Electronic Physician: Excellence in Constructive Peer Review* [1.97]. This article provides an account of so-called "bogus" indicators such as the Universal Impact Factor (UIF), Global Impact Factor (GIF), and Citefactor. The article describes the threat of these indicators to reputable indicators such as Thomson Reuters *and* the Index Copernicus metric value (ICV). However, the ICV, which is prominently displayed on the website of the *Electronic Physician*, is itself under scrutiny for its association with predatory journals [1.98]. Therefore, this article seems to provide much the same function as the pop-up box of the International Impact Factor Services: It is a classic redirect technique, wherein the service attempts to legitimatize their own activities by delegitimizing others.

One of the biggest concerns with these products is the lack of transparency in the compilation of the indicators. The *Global Impact Factor* obliquely combines some form of peer review with the number of papers published [1.99]. Journals of the "Academy of IRMBR International Research in Management and Business Realities"—contained in Beall's list—rely on GoogleScholar to generate indicators [1.100], which seems a common approach for these fake JIFs. While one could argue that many of these indicators are legitimate competitors, rather than exploitative knock-offs, the mimicry of the names and acronyms as well as the cost structure begs caution. For example, the Global Impact Factor provides their indicator for an annual fee of $40 [1.99] and *International Scientific Indexing* charges $100–130 per journal for the indicator and indexation on their platform [1.101]. While the deceptive character of these sites might be apparent to many scholars, some have chosen to take a more neutral stance. For instance, a US university library guide on journal indicators lists these indicators alongside the JIF and other established indicators [1.102]. Other libraries have taken a more direct stance, urging their audience caution with these indicators and predatory publishers [1.103].

1.5 What Are the Alternatives?

Knock-off indicators abound, but there are also several other indicators that have emerged as complementary to or competitive with the JIF. This section examines four of the most established: the group of Eigenfactor Metrics, Source Normalized Impact per Paper (SNIP), CiteScore, and SCImago Journal Rank (SJR).

The Eigenfactor Metrics were introduced in 2010 as a new approach for ranking journals [1.104]. The metrics include two related indicators—the Eigenfactor Score and Article Influence Score (AIS)—both based on the Eigenfactor algorithm, which leverages the citation network to identify and weight citations from central journals. The underlying algorithm is derived from *Phillip Bonacich*'s [1.105] eigenvector centrality, which has been employed across several domains, most notably as the foundation for Google's PageRank algorithm. The Eigenfactor Score depicts the *total value* of a journal and is thus size-dependent—as the size of the journal increases, so too will the Eigenfactor Score. The Article Influence Score, however, measures the average influence of articles in the journal, and is therefore more comparable to the JIF. However, there are several important differences: the AIS is calculated over a five-year (rather than two-year) time window, excludes self-citations, and uses weighted citations. Like the JIF, both indicators rely on Web of Science (WoS) data and were added to the JCR in 2009. As such, they represent a supplement to the JCR portfolio, rather than direct competition.

Scopus—the largest competitor to Web of Science—also has several associated journal indicators. The Source Normalized Impact per Paper (SNIP) indicator was proposed in 2009 by *Henk F. Moed*, then at the Center for Science and Technology Studies (CWTS) of Leiden University [1.106] and later revised by *Waltman* and colleagues [1.107]. As discussed, one of the central interpretive critiques of the JIF is the inability to make cross-disciplinary comparisons. SNIP was developed to account for the different *citation potential* among fields. Rather than using an a priori journal-based classification, fields are defined according to the set of citing papers. In this way, the indicator is based on *contextual*, rather than absolute, citation impact. Furthermore, SNIP serves to address another limitation of the JIF: by focusing on the set of citing papers, there is no concern about the asymmetries created by noncitable items. However, like the JIF, self-citations are included, which can lead to distortions in extreme cases. Furthermore, SNIP tends to be higher in journals with a large proportion of review articles, which causes additional bias. SNIP uses a three-year citation window—one year more than the JIF, but two less than the Article Influence Score.

Another indicator contained in "the Scopus basket of journal metrics" [1.108] is the SCImago Journal Rank (SJR), which was developed and continues to be updated by the SCImago research group at the University of Granada [1.109]. Like the Eigenfactor Score, the SJR employs Bonacich's eigenvector centrality to calculate the prestige of a journal, weighting the links according to the closeness of co-citation relationships (on the basis of citable documents). The current version of the indicator uses a three-year window, in keeping with the other Scopus journal indicators [1.110]. Furthermore, several heuristics are applied to circumvent gaming and distortions: in generating the prestige of the journal, there are thresholds on how much a single journal and the journal itself can provide—protecting against citation cartels and self-citations—and prestige is calculated on the basis of proportions rather than number of citable documents, to control for size and the dynamicity of the database.

In 2016, Elsevier released a new journal impact indicator with the name CiteScore [1.108]. The indicator is obtained by averaging, for a given journal, the number of citations received in a single year by papers it published during the preceding three years. The appeal is the simplicity—it is merely an average of citations received for all document types, which removes concerns about asymmetries between cited and citing items. However, the inclusion of all document types shifts the bias in another direction. While journals with a high proportion of noncitable items (e.g., editorials, news items) tend to fare well in the JIF, they are ranked lower in CiteScore. Critics of CiteScore have noted that this favorably biases Elsevier's own journals, which tend to publish a lower proportion of front matter than other journals (such as Nature's journals) [1.111]. Broader concerns have also been raised about the conflict of interest inherent in vertically integrated companies: There is considerable concern about the construction of indicators within a company that also publishes, indexes, and provides analytic services for journals [1.112]. The increasing monopoly of Elsevier in this space has caused some to question the neutrality of the indicator.

However, none of these indicators have managed to displace the JIF's role in the scientific system. The Eigenfactor Metrics are included in the JCR, but have not gained the marketing appeal of the JIF, and the Scopus indicators have also not gained widespread traction after nearly a decade of existence. Part of this is the appeal of standardization: scholars working in research evaluation (whether hiring, promoting, or granting) have internalized the value of the JIF. Despite the well-known technical and interpretive concerns, the JIF remains the standard journal indicator.

1.6 The Future of Journal Impact Indicators

Building upon both original data and a review of the literature, this chapter provides a background for the creation of the JIF, an overview of its limitations, and a discussion of some of the most documented adverse effects. Several of the technical critiques can be or already are addressed by Clarivate. For instance, asymmetries between the numerator and denominator could be controlled by more careful analysis and cleaning of the data. Journal self-citations account for a minority of citations and can (and are already) flagged when excessive. The two-year JIF could be removed, in favor of a JIF with a longer citation window—which is already provided in more recent editions of the JCR. However, rather than replacing the original JIF with new indicators, these alternatives have merely been added to the JCR. This multiplicity of indicators is problematic from the perspective of standardization. When every researcher, administrator, evaluator, and policy-maker is constructing tailor-made indicators, the indicators lose their central function—to communicate globally and across disciplines in a standard fashion [1.112]. Of course, bibliometrics is not alone in dedication to an imperfect indicator. For example, despite heavy criticism and the creation of alternative indicators [1.113, 114], the body mass index remains, as per the World Health Organization, the standard for the measurement of obesity.

However, some of the most disconcerting aspects are not purely technical, but rather due to the misapplication of the indicator. For example, one common technical concern is the skewness of citation distributions. Given that less than a third of articles are likely to achieve the citation value of the JIF, the indicator is misleading for application at the individual paper level. Because of the skewness of citation distributions and the declining predictive power of the JIF, it is widely acknowledged that the indicator should not be used to evaluate individual articles or scholars (though there remains debate on this issue [1.115]). Furthermore, the lack of normalization by discipline and the continual inflation of the indicator over time means that the JIF can only be used to rank contemporary journals within the same discipline.

It is also clear that it is not the indicator, but rather the application of the indicator that is causing systemic disruptions in science. Several of the adverse effects observed are not directly linked to JIF; rather, they are linked to the research evaluation system and, more specifically, to journals as vectors of scientific capital. In other words, the JIF has become synonymous with academic capital, and despite well-publicized criticisms [1.90], it remains central to research evaluation. It would, of course, be naïve to assume that, in a pre-JIF era, there was no relationship between economic and scientific capital. Journals have long served at the heart of the race for scientific discovery: the certification and dissemination of knowledge allowed scholars to make priority claims, the traditional building blocks of scientific reputation [1.116]. However, the direct relationship between cash rewards and JIF is a gross perversion of the reward system in which economic incentives become the main objective of publishing. It is clear that measure has become the target [1.117], as evident by the explicit manipulations within the system and the gross goal displacement in favor of high-impact journals, whereby there is a prioritization of metrics over ethics [1.57, 87].

When he published the first iteration of the JCR, *Garfield* hoped that it could

> prove itself indispensable to people who cannot rely on economic criteria alone in making basic decisions about journals, since the law of supply and demand is not always allowed to prevail [1.5, p. 1].

The JIF became more than that: in many ways, it has become itself an economic item, capitalizing upon academic capital and the need for its measurement. As such, it has been grossly misapplied to make decisions about papers and authors, rather than journals, and caused distortions within the scholarly system. And while *Garfield* foresaw the use of the JIF for research evaluation, he also formulated recommendations for its proper use in his introduction of the first JCR [1.4, p. 1]:

> Like any other tool, the JCR cannot be used indiscriminately. It is a source of highly valuable information, but that information must be used within a total framework proper to the decision to be made, the hypothesis to be examined, and rarely in isolation without consideration of other factors, objective and subjective.

Among these subjective factors, Garfield noted the reputation of the author, the controversial nature of the subject, the circulation and cost of the journal, and the degree to which the work is accessible. Garfield cautioned against comparing citation rates for journals in different disciplines and noted the biases in accounting for journals which do not use the Roman alphabet. While those factors remain quite relevant today, it seems they have been forgotten along the way. Moreover, since Garfield made these recommendations, English has become the lingua franca of research [1.118], which has led to a decline of the relative importance of non-English journals in many disciplines and, thus, reinforced the Web of Science—and, by extension, the JCR—as a measurement tool.

Despite these well-documented limitations and consequences, the JIF will likely remain part of the research ecosystem and as long as journals remain the primary mechanism for diffusing new knowledge, their reputation—as established by JIF or an alternative—will remain a marker of capital. It is essential, therefore, that actors within this system are provided with the means to interpret and apply the indicators responsibly, in full awareness of the consequences [1.12, 119]. Perhaps more importantly, the scientific community must collectively ask: is the use of the journal impact factor good for science?

References

1.1 E. Garfield, I.H. Sher: New factors in the evaluation of scientific literature through citation indexing, Am. Doc. **14**(3), 195–201 (1963)

1.2 E. Garfield: Citation analysis as a tool in journal evaluation, Science **178**(4060), 471–479 (1972)

1.3 É. Archambault, V. Larivière: History of the journal impact factor: contingencies and consequences, Scientometrics **79**(3), 639–653 (2009)

1.4 E. Garfield: Introduction I. Journals, References, and Citations. In: *Journal Citation Reports ® A Bibliometric Analysis of References Processed for the 1974 Science Citation Index*, Science Citation Index, Vol. 9, ed. by E. Garfield (1976)

1.5 E. Garfield: Introduction II. Why the Journal Citation Reports? In: *Journal Citation Reports® A Bibliometric Analysis of References Processed for the 1974 Science Citation Index*, Science Citation Index, Vol. 9, ed. by E. Garfield (1976)

1.6 P.L.K. Gross, E.M. Gross: College libraries and chemical education, Science **66**(1713), 385–389 (1927)

1.7 E. Garfield: Preface. In: *Journal Citation Reports® A Bibliometric Analysis of References Processed for the 1974 Science Citation Index*, Science Citation Index, Vol. 9, ed. by E. Garfield (Institute for Scientific Information, Philadelphia 1976)

1.8 D. de Solla Price: *Little Science, Big Science* (Columbia Univ. Press, New York 1963)

1.9 R. Monastersky: The number that's devouring science, The Chronicle of Higher Education, http://www.chronicle.com/article/The-Number-Thats-Devouring/26481 (2005)

1.10 R. Müller, S. De Rijcke: Thinking with indicators. Exploring the epistemic impacts of academic performance indicators in the life sciences, Res. Eval. (2017), https://doi.org/10.1093/reseval/rvx023

1.11 M.E. McVeigh, S.J. Mann: The journal impact factor denominator. Defining citable (counted) items, J. Am. Med. Assoc. **302**(10), 1107–1109 (2009)

1.12 D. Hicks, P. Wouters, L. Waltman, S. De Rijcke, I. Rafols: The Leiden Manifesto for research metrics, Nature **520**(7548), 429–431 (2015)

1.13 F. Anseel, W. Duyck, W. De Baene, M. Brysbaert: Journal impact factors and self-citations: Implications for psychology journals, Am. Psychol. **59**, 49–51 (2004)

1.14 M. Rossner, H. Van Epps, E. Hill: Show me the data, J. Cell Biol. **179**(6), 1091–1092 (2007)

1.15 S. Curry: Sick of Impact Factors, Reciprocal Space, http://occamstypewriter.org/scurry/2012/08/13/sick-of-impact-factors/ (2012)

1.16 P. Campbell: Escape from the impact factor, Ethics Sci. Environ. Polit. **8**(1), 5–7 (2008)

1.17 S.C. Hubbard, M.E. McVeigh: Casting a wide net: the journal impact factor numerator, Learn. Publ. **24**, 133–137 (2011)

1.18 T. Kiesslich, S.B. Weineck, D. Koelblinger: Reasons for journal impact factor changes: influence of changing source items, PLOS ONE **11**(4), e0154199 (2016)

1.19 H.F. Moed, T.N. van Leeuwen: Improving the accuracy of Institute for Scientific Information's journal impact factors, J. Am. Soc. Inf. Sci. **46**(6), 461–467 (1995)

1.20 H.F. Moed, T.N. van Leeuwen: Impact factors can mislead, Nature **381**(6579), 186 (1995)

1.21 D.N. Arnold, K.K. Fowler: Nefarious numbers, Notices AMS **58**(3), 434–437 (2011)
1.22 J. Reedijk, H.F. Moed: Is the impact of journal impact factors decreasing?, J. Doc. **64**(2), 183–192 (2008)
1.23 B.R. Martin: Whither research integrity? Plagiarism, self-plagiarism, and coercive citation in the age of research assessment, Res. Policy **42**, 1005–1014 (2013)
1.24 R. Rousseau: Temporal differences in self-citation rates of scientific journals, Scientometrics **44**, 521–531 (1999)
1.25 M.E. McVeigh: Journal self-citation in the Journal Citation Reports–Science Edition (2002), http://wokinfo.com/essays/journal-self-citation-jcr/ (2002)
1.26 T.E. Nisonger: Use of Journal Citation Reports for serials management in research libraries: An investigation of the effect of self-citation on journal rankings in library and information science and genetics, Coll. Res. Libr. **61**, 263–275 (2000)
1.27 T. Opthof: Inflation of impact factors by journal self-citation in cardiovascular science, Neth. Heart J. **21**(4), 163–165 (2013)
1.28 K. Hamilton: *Subfield and Level Classification of Journals*, CHI Report, Vol. 2012-R (CHI Research, Cherry Hill 2003)
1.29 W. Glänzel, H.F. Moed: Journal impact measures in bibliometric research, Scientometrics **53**(2), 171–193 (2002)
1.30 A.J. Lotka: The frequency distribution of scientific productivity, J. Wash. Acad. Sci. **16**(12), 317–323 (1926)
1.31 P.O. Seglen: The skewness of science, J. Am. Soc. Inf. Sci. **43**(9), 628–638 (1992)
1.32 V. Larivière, V. Kiermer, C.J. MacCallum, M. McNutt, M. Patterson, B. Pulverer, S. Swaminathan, S. Taylor, S. Curry: A simple proposal for the publication of journal citation distributions, Biorxiv, p.062109, http://www.biorxiv.org/content/early/2016/07/05/062109 (2016)
1.33 N. Sombatsompop, T. Markpin, N. Premkamolnetr: A modified method for calculating the impact factors of journals in ISI Journal Citation Reports: Polymer science category in 1997–2001, Scientometrics **60**, 217–235 (2004)
1.34 R. Rousseau: Median and percentile impact factors: A set of new indicators, Scientometrics **63**(3), 431–441 (2005)
1.35 V. Larivière, Y. Gingras: The impact factor's Matthew effect: A natural experiment in bibliometrics, J. Assoc. Inf. Sci. Technol. **61**(2), 424–427 (2010)
1.36 G.A. Lozano, V. Larivière, Y. Gingras: The weakening relationship between the impact factor and papers' citations in the digital age, J. Assoc. Inf. Sci. Technol. **63**(11), 2140–2145 (2012)
1.37 A. Acharya, A. Verstak, H. Suzuki, S. Henderson, M. Iakhiaev, C.C.Y. Lin, N. Shetty: Rise of the rest: The growing impact of nonelite journals, arXiv:1410.2217 (2014)
1.38 V. Larivière, Y. Gingras, É. Archambault: The decline in the concentration of citations, 1900-2007, J. Assoc. Inf. Sci. Technol. **60**(4), 858–862 (2009)
1.39 V. Larivière, É. Archambault, Y. Gingras: Long-term variations in the aging of scientific literature: From exponential growth to steady-state science (1900–2004), J. Assoc. Inf. Sci. Technol. **59**(2), 288–296 (2008)
1.40 R.E. Lucas: Econometric policy evaluation: A critique. In: *The Phillips Curve and Labor Markets*, Carnegie-Rochester Conference Series on Public Policy, Vol. 1, ed. by K. Brunner, A. Meltzer (Elsevier, New York 1976) pp. 19–46
1.41 A. Schubert, T. Braun: Relative indicators and relational charts for comparative assessment of publication output and citation impact, Scientometrics **9**(5–6), 281–291 (1986)
1.42 H.F. Moed, R.E. De Bruin, T.N. van Leeuwen: New bibliometric tools for the assessment of national research performance: database description, overview of indicators and first applications, Scientometrics **33**(3), 381–422 (1995)
1.43 R. Bolli: The impact factor of Circulation Research reaches a new high, Circulation **121**(3), 199 (2017)
1.44 M. Govendir: Veterinary Quarterly 2015 Impact Factor increases to 1.047, Vet. Q. **36**(3), 121 (2016)
1.45 M. Simren, G. Mawe, J. Galligan, S. Bradesi, A. Bredenoord: Impact factor increases to its highest level ever, Neurogastroenterol. Motil. **27**(8), 1051 (2015)
1.46 Elsevier: Elsevier announces significant increases in impact factors, https://www.elsevier.com/about/press-releases/research-and-journals/elsevier-announces-significant-increases-in-impact-factors (2007)
1.47 Springer: Impact factors of Springer and Palgrave Macmillan journals announced: increased number of journals in JCR, https://www.springer.com/gp/about-springer/media/press-releases/corporate/impact-factors-of-springer-and-palgrave-macmillan-journals-announced-increased-number-of-journals-in-jcr/10286018 (2016)
1.48 Wiley.: Wiley journals increase Impact Factors in 2015 Journal Citation Reports, http://www.wiley.com/WileyCDA/PressRelease/pressReleaseId-126482.html?campaign=jrpm (2016)
1.49 M. Sosteric: Endowing mediocrity: neoliberalism, information technology, and the decline of radical pedagogy, Radic. Pedagog. **1**(1) (1999)
1.50 D.T. Campbell: Assessing the impact of planned social change, Eval. Program Plan. **2**, 67–90 (1979)
1.51 C. Goodhart: Monetary relationships: a view from Threadneedle Street. In: *Papers of Monetary Economics*, Vol. 1 (Reserve Bank of Australia, Sydney 1975)
1.52 C. Goodhart: Problems of monetary management: the UK experience. In: *Papers of Monetary Economics*, Vol. 1 (Reserve Bank of Australia, Sydney 1975)
1.53 B. Cronin, C.R. Sugimoto (Eds.): *Scholary Metrics Under the Microscope: Citation Analysis and Academic Auditing* (InfoToday/ASIST, Medford 2015), p. 976

1.54 R. Smith: Journal accused of manipulating impact factor, BMJ **312**, 461 (1997)
1.55 T.F. Frandsen: Journal self-citations—Analysing the JIF mechanism, J. Infometr. **1**, 47–58 (2007)
1.56 R. Noorden: Record number of journals banned for boosting impact factor with self-citations, Nature News Blog, http://blogs.nature.com/news/2012/06/record-number-of-journals-banned-for-boosting-impact-factor-with-self-citations.html (2012)
1.57 G. Franck: Scientific communication—a vanity fair?, Science **286**, 53–56 (1999)
1.58 A.W. Wilhite, E.A. Fong: Coercive citation in academic publishing, Science **335**, 542–543 (2012)
1.59 E.A. Fong, A.W. Wilhite: Authorship and citation manipulation in academic research, PLOS ONE **12**(12), e01897394 (2017)
1.60 B. Cronin: Do me a favor, J. Am. Soc. Inf. Sci. Technol. **63**(7), 1281 (2012)
1.61 P. Davis: Citation cartel or editor gone rogue?, The Scholarly Kitchen, https://scholarlykitchen.sspnet.org/2017/03/09/citation-cartel-or-editor-gone-rogue/ (2017)
1.62 P. Davis: The emergence of a citation cartel, The Scholarly Kitchen, https://scholarlykitchen.sspnet.org/2012/04/10/emergence-of-a-citation-cartel/ (2012)
1.63 P. Mongeon, L. Waltman, S. de Rijcke: What do we know about journal citation cartels? A call for information, CWTS Blog, https://www.cwts.nl/blog?article=n-q2w2b4 (2016)
1.64 P. Heneberg: From excessive journal self-cites to citation stacking: Analysis of journal self-citation kinetics in search for journals, which boost their scientometric indicators, PLoS ONE **11**(4), e0153730 (2016)
1.65 Clarivate Analytics: Title suppression from Journal Citation Reports, http://wokinfo.com/media/pdf/jcr-suppression.pdf (2017)
1.66 Thomson Reuters: Title suppressions, http://ipscience-help.thomsonreuters.com/incitesLiveJCR/JCRGroup/titleSuppressions.html (2016)
1.67 P. Davis: Reverse Engineering JCR's Self-Citation and Citation Stacking Thresholds, The Scholarly Kitchen, https://scholarlykitchen.sspnet.org/2017/06/05/reverse-engineering-jcrs-self-citation-citation-stacking-thresholds/ (2017)
1.68 S. Hubbard: Journal suppressions in the 2015 JCR data—why so few?, http://clarivate.com/journal-suppressions-in-the-2015-jcr-data-why-so-few/ (2016)
1.69 C. Chorus, L. Waltman: A large-scale analysis of impact factor biased journal self-citations, PLoS ONE **11**(8), e0161021 (2016)
1.70 W. Quan, B.-C. Chen, F. Shu: Publish or poverish: An investigation of the monetary reward system of science in China, Aslib J. Inf. Manag. **69**(4), 1999–2016 (2017)
1.71 L. Butler: Explaining Australia's increased share of ISI publications: the effects of a funding formula based on publication counts, Res. Policy **32**, 143–155 (2003)
1.72 J. Qiu: Publish or perish in China, Nature **463**(7278), 142–143 (2010)
1.73 M. Hvistendahl: China's publication bazaar, Science **342**(6162), 1035–1039 (2013)
1.74 C. Franzoni, G. Scellato, P. Stephan: Science policy. Changing incentives to publish, Science **333**(6043), 702–703 (2011)
1.75 V. Calcagno, E. Demoinet, K. Gollner, L. Guidi, D. Ruths, C. de Mazancourt: Flows of research manuscripts among scientific journals reveal hidden submission patterns, Science **338**, 1065–1069 (2012)
1.76 R.K. Merton: *The Sociology of Science: Theoretical and Empirical Investigation* (Univ. Chicago Press, Chicago 1973)
1.77 B. Jin, R. Rousseau: Evaluation of research performance and scientometric indicators in China. In: *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems*, ed. by H.F. Moed, W. Glänzel, U. Schmoch (Springer, Dordrecht 2004) pp. 497–514
1.78 E. Garfield: English – An international language for science?. In: *Essays of an Information Scientist*, Vol. 1 (ISI, Philadelphia 1977) p. 19–20
1.79 D. Cryanoski: Strong medicine for China's journals, Nature, http://www.nature.com/news/2010/100915/full/467261a.html (2010)
1.80 J. Shao, H. Shen: The outflow of academic papers from China: why is it happening and can it be stemmed?, Learn. Publ. **24**, 95–97 (2011)
1.81 V. Larivière: De l'importance des revues de recherche nationales, Découvrir, la revue de l'ACFAS, http://www.acfas.ca/publications/decouvrir/2014/09/l-importance-revues-recherche-nationales (2014)
1.82 B. Alberts: Impact factor distortions, Science **340**, 787 (2013)
1.83 R. Noorden: Metrics: A profusion of measures, Nature **465**, 864–866 (2010)
1.84 H.F. Moed: The impact-factors debate: the ISI's uses and limits, Nature **415**, 731–732 (2002)
1.85 A. Rushforth, S. de Rijcke: Accounting for impact? The journal impact factor and the making of biomedical research in the Netherlands, Minerva **53**(2), 117–139 (2015)
1.86 R.A. Brumback: "3...2...1... Impact [Factor]: Target [Academic Career] Destroyed!": Just another statistical casualty, J. Child Neurol. **27**(12), 1565–1576 (2012)
1.87 M. Osterloh, B.S. Frey: Ranking Games, Eval. Rev. **39**(1), 102–129 (2014)
1.88 H.W. Holub, G. Tappeiner, V. Eberharter: The iron law of important articles, South. Econ. J. **58**, 317–328 (1991)
1.89 S. Shibayama, Y. Baba: Impact-oriented science policies and scientific publication practices: The case of life sciences in Japan, Res. Policy **44**, 936–950 (2015)

1.90 ASCB: San Francisco Declaration on Research Assessment, http://www.ascb.org/dora/ (2012)
1.91 NHMRC: NHMRC removes journal impact factors from peer review of individual research grant and fellowship applications, https://www.nhmrc.gov.au/_files_nhmrc/file/grants/peer/impact%20factors%20in%20peer%20review.pdf (2010)
1.92 P. Doherty: How important is a journal's impact factor? Nobel Prize Inspiration Initiative, http://www.nobelprizeii.org/videos/important-journals-impact-factor/ (2015)
1.93 I. Sample: Nobel winner declares boycott of top science journals, The Guardian, https://www.theguardian.com/science/2013/dec/09/nobel-winner-boycott-science-journals (2013)
1.94 J. Beall: Beall's list of predatory journals and publishers, http://beallslist.weebly.com/misleading-metrics.html (2017)
1.95 Global Institute for Scientific Information: Journal Impact Factor (JIF), http://www.jifactor.com/index.asp (2017)
1.96 International Impact Factor Services: Homepage, http://impactfactorservice.com/ (2017)
1.97 M. Jalalian: The story of fake impact factor companies and how we detected them, Electron Phys. **7**(2), 1069 (2015)
1.98 J. Beall: Index Copernicus has no value. Scholarly Open Access, https://web.archive.org/web/20140403010813/http://scholarlyoa.com/2013/11/21/index-copernicus-has-no-value/ (2013)
1.99 Global Impact Factor: Evaluation method, http://globalimpactfactor.com/evaluatiion-method/ (2017)
1.100 Academy of IRMBR: International Review of Social Sciences (IRSS), https://scholar.google.com.my/citations?user=k7ZMMxIAAAAJ&hl=en (2017)
1.101 International Scientific Indexing.: Payments, http://isindexing.com/isi/payment.php (2017)
1.102 Cal State East Bay: Analyzing scholarly research: Bibliometrics, altmetrics, journal impact factors, citation analysis: Journal Impact Factors, http://library.csueastbay.edu/analyzingscholarlyresearch/journalimpactfactors (2017)
1.103 George Washington University: Himmelfarb Health Sciences Library: Scholarly Publishing: Predatory Publishing, http://libguides.gwumc.edu/scholarlypub/predatory_publishing (2017)
1.104 J.D. West, T.C. Bergstrom, C.T. Bergstrom: The Eigenfactor™ metrics: a network approach to assessing scholarly journals, Coll. Res. Libr. **71**, 236–244 (2010)
1.105 P. Bonacich: Factoring and weighting approaches to status scores and clique identification, J. Math. Sociol. **2**(1), 113–120 (1972)
1.106 H.F. Moed: Measuring contextual citation impact of scientific journals, J. Informetr. **4**, 265–277 (2010)
1.107 L. Waltman, N.J. van Eck, T.N. van Leeuwen, M.S. Visser: Some modifications to the SNIP journal impact indicator, J. Informetr. **7**(2), 272–285 (2013)
1.108 H. Zijlstra, R. McCullough: CiteScore: a new metric to help you track journal performance and make decisions, Elsevier, https://www.elsevier.com/editors-update/story/journal-metrics/citescore-a-new-metric-to-help-you-choose-the-right-journal (2016)
1.109 B. González-Pereira, V.P. Guerrero-Bote, F. Moya-Anegón: A new approach to the metric of journals scientific prestige: The SJR indicator, J. Informetr. **4**(3), 379–391 (2010)
1.110 V.P. Guerrero-Bote, F. Moya-Anegón: A further step forward in measuring journals' scientific prestige: The SJR2 indicator, J. Informetr. **6**, 674–688 (2012)
1.111 C.T. Bergstrom, J. West: Comparing Impact Factor and Scopus CiteScore, Eigenfactor.org, http://eigenfactor.org/projects/posts/citescore.php (2016)
1.112 C.R. Sugimoto, V. Larivière: *Measuring Research: What Everyone Needs to Know* (Oxford Univ. Press, Oxford 2018)
1.113 A. Bagust, T. Walley: An alternative to body mass index for standardizing body weight for stature, Q. J. Med. **93**(9), 589–596 (2000)
1.114 M.K. Lebiedowska, K.E. Alter, S.J. Stanhope: Human body shape index based on an experimentally derived model of human growth, J. Pediatr. **152**(1), 45–49 (2008)
1.115 L. Waltman, V.A. Traag: Use of the journal impact factor for assessing individual articles need not be wrong. arXiv:1703.02334 (2017)
1.116 P. Dasgupta, P.A. David: Toward a new economics of science, Res. Policy **23**, 487–521 (1994)
1.117 M. Strathern: Improving ratings: audit in the British university system, Eur. Rev. **5**(3), 305–321 (1997)
1.118 S.L. Montgomery: *Does science need a global language? English and the future of research* (Univ. Chicago Press, Chicago 2013)
1.119 P. Lawrence: The politics of publication, Nature **422**, 259–261 (2003)

Vincent Larivière

École de bibliothéconomie et des sciences de l'information
Université de Montréal
Montréal, Canada
vincent.lariviere@umontreal.ca

Vincent Larivière is Associate Professor of Information Science at the École de bibliothéconomie et des sciences de l'information of the Université de Montréal, where he teaches research methods and bibliometrics. He is also the Scientific Director of the Érudit journal platform, Associate Scientific Director of the Observatoire des sciences et des technologies and a Regular Member of the Centre interuniversitaire de recherche sur la science et la technologie.

Cassidy R. Sugimoto

School of Informatics and Computing
Indiana University Bloomington
Bloomington, IN, USA
sugimoto@indiana.edu

Cassidy R. Sugimoto is Associate Professor at the School of Informatics, Computing and Engineering of Indiana University, where she researches within the domain of Scholarly Communication and Scientometrics, examining the formal and informal ways in which knowledge producers consume and disseminate scholarship. Sugimoto has an undergraduate degree in Music Performance, MS in Library Science, and PhD in Information and Library Science from the University of North Carolina.

2. Bibliometric Delineation of Scientific Fields

Michel Zitt, Alain Lelu, Martine Cadot, Guillaume Cabanac

Delineation of scientific domains (fields, areas of science) is a preliminary task in bibliometric studies at the mesolevel, far from straightforward in domains with high multidisciplinarity, variety, and instability. The Sect. 2.2 shows the connection of the delineation problem to the question of disciplines versus invisible colleges, through three combinable models: ready-made classifications of science, classical information-retrieval searches, mapping and clustering. They differ in the role and modalities of supervision. The Sect. 2.3 sketches various bibliometric techniques against the background of information retrieval (IR), data analysis, and network theory, showing both their power and their limitations in delineation processes. The role and modalities of supervision are emphasized. The Sect. 2.4 addresses the comparison and combination of bibliometric networks (actors, texts, citations) and the various ways to hybridize. In the Sect. 2.5, typical protocols and further questions are proposed.

2.1 Shaping the Landscape of Scientific Fields

Collecting literature that is both relevant and specific to a domain is a preliminary step of many scientometric studies: description of strategic fields such as nanosciences, genomics and proteomics, environmental sciences; research monitoring and international benchmarks; science community analyses. Although our focus here is on the intermediate levels, informally described in such terms as areas, specialties, subfields, fields, subdisciplines… this subject is connected to general science classification and, at the other end of the range, to narrow topic search.

In Sect. 2.2 we place delineation at the crossroads of two concepts: the first one is *disciplinarity* (what is a scientific discipline?), which crystallizes various dimensions of scientific activity in epistemology and sociology. The second one is *invisible colleges* in resonance with the core of bibliometrics, the study of networks created explicitly or implicitly by publishing actors. From this point of view, domains of science can be viewed as a generalized form of invisible colleges, sometimes in the form of relatively dense and segregated areas—at some scale. In other cases however, the structure is less clear and bounded, with high levels of both internal diversity and external connections and overlaps. Given a target domain, its expected diversity, interdisciplinarity, and instability are challenging issues. We outline the main approaches to delineation: external formalized resources, such as science classifications; ad hoc information retrieval (IR) search; network exploration resources (clustering–mapping).

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_2

Section 2.3 is devoted to the main approaches in domain delineation, IR search, and science clustering–mapping, when off-the-shelf classifications are not sufficient. Both take root in the information networks of science, but start from different vantage points, with some simplification: ex ante heavy supervision for IR search, typically with bottom-up ad hoc queries; ex post supervision for bibliometric mapping, with top-down pruning. In difficult cases, these approaches appear complementary, often within multistep protocols. As a result of the complex structure and massive overlaps of aspects of science, of the multiple bibliometric networks involved, of the multiple points of view, the frontiers are far from unique at a given scale of observation. The experts' supervision process is a key element. Its organization depends on the studies' context and demand, to reach decisions through confrontation and negotiation, especially in high-stakes contexts. Beforehand, we shall briefly address the toolbox of data analysis methods for clustering–mapping purposes.

Section 2.4 focuses on the multinetwork approach for delineation tasks, stemming from pragmatic practices of information retrieval (IR) and bibliometrics. The main networks are actor's graphs and other relations connected with invisible colleges based on documents and their main attributes, texts, and citations. Other scientometric networks (teaching, funding, science social networks, etc.) offer potential resources. The hybridization covers a wide scope of forms. There is a strong indication that multinetwork methods improve IR performance and offer a richer substance to experts'/users' discussions.

2.2 Context

2.2.1 Background: Disciplinarity and Invisible Colleges

Generally speaking there is no ground truth basis for defining scientific domains. Given a target domain, assigned by sponsors in broad and sometimes fuzzy terms, delineation is the first stage of a bibliometric study. It is tantamount to a rule of decision involving sponsors/stakeholders, scientists/experts, and bibliometricians on extraction of the relevant literature. Delineation also matters as research communities are an object of science sociology as well as a playground for network theoreticians.

The delineation of scientific domains should be understood in the context of the structure of science and scientific communities, especially through the game between diversity, source of speciation, and interdisciplinarity drive towards reunification. Disciplinarity and *invisible* colleges are two concepts from the sociology of science that symbolize two kinds of communities, the first one more formal and institutional, the second one constructed on informal linkages made visible by bibliometric analysis of science networks. The tradition of epistemology has contributed to highlight the specificity of science by contrast to other conceptions of knowledge. *Auguste Comte* proposed the first modern classification of science and at the same time condemned the drift of specialization [2.1], considered a threat to a global understanding of positive science. In reaction both to epistemology and normative Mertonian tradition [2.2, 3], *Kuhn* emphasized the role of central paradigms in disciplines at some point of their evolution [2.4]. The post-Kuhnian social constructivism proceeded along two lines—at times conflicting [2.5]—of relativist thinking: the *strong programme* (see *Barnes* et al. [2.6]) and the no less radical actor–network theory (ANT). The first one was initiated by Barnes and *Bloor* [2.7] and flourished in the science studies movement [2.8, 9]. The ANT also borrowed from Serres (*translation* concept [2.10]) and from the poststructuralist French theory (Foucault, Derrida, Bourdieu, Baudrillard), see [2.11–13]. These schools of thought emphasize disciplinarity rather than unity. *Lenoir* notes that [2.14, pp. 71–72, 82]:

> A major consequence of [social constructivism] has been to foreground the heterogeneity of science. [... Disciplines are] crucial sites where the skills [originating in labs] are assembled and political institutions that demarcate areas of academic territory, allocate privileges and responsibilities of expertise, and structure claims on resources.

Bourdieu stressed the importance of personal relationship and *shared habitus*. Disciplines exhibit both a strong intellectual structure and a strong organization. The institutional framework, with, in most countries, an integration of research and higher education systems, ensures evaluation and career management. Some communities coin their own jargon, amongst signs of differentiation, and norms and patterns. Potentially, all dimensions of research activity (paradigms and theories, classes of problems, methodology and tools, shared vocabulary, corroboration protocols, construction of scientific facts and interpretation) appear as discipline-informed, with particular tensions between superdisciplines, natural sciences and social sciences and humanities. Scientists discuss, within their own disciplines, the subfield breakdown and the structuring role of particular dimensions, for example research objects in microbiology, versus integration drive [2.15, 16].

The endless process of specialization and speciation in science, erecting barriers to the mutual understanding of scientists, is partly counteracted by interdisciplinary linkages which maintain and create solidarity between neighbor or remote areas of research. *Piaget* [2.17] coined the term transdisciplinarity as the new paradigm re-engaging with unity of science. A few rearrangements of large magnitude, such as the movement of convergence between nanosciences, biomedicine, information, and cognitive sciences and technologies (NBIC, Nanotechnology, biotechnology, information technology, and cognitive science; concept coined by NSF (National Science Foundation) in 2002), tend to reunite distant areas or at least create active zones of overlap.

In contrast with disciplinarity, the concept of *invisible college* in its modern acceptation, popularized by *Price* and *Beaver* [2.18] and *Crane* [2.19], chiefly refers to informal communication networks, personal relationship, and possibly interdisciplinary scope. These direct linkages tend to limit the size of the colleges, although no precise limit can be given. Science studies devote a large literature to those informal groups, which exemplify how networks of actors operate at various levels of science [2.20, 21].

Although more formal expressions emerge from the self-organization of those microsocieties (workshops, conferences, journals), the invisible colleges do not claim the relative stability and the social organization of disciplines. The various communication phenomena of the colleges are revealed by sociological studies or, more superficially but systematically, by analysis of bibliometric networks such as coauthorship, text relations and citations. The *bibliometric hypothesis* assumes that the latter process mirrors essential aspects of science: the traceable publication activity, in a broad sense, expresses the collective behavior of scientific communities in most relevant aspects (contents and certification, production and structure of knowledge, diffusion and reward, cooperation and self-organization). It does not follow that bibliometrics can easily operationalize all hypotheses [2.22]. Affiliations can, in the background, connect to the layers of academic institutions or corporate entities. Mentions to funding bodies are increasingly required in articles reporting grant-supported works. These relations, however, as well as personal interactions, generally require extrabibliometric information. Variants of the invisible colleges in sociology of science are known as epistemic communities, involving scientists and experts with shared convictions and norms [2.8, 23] and community of practice [2.24]. The mix of behavior, stakes and power games, in the interaction of virtual colleges and institutions, remains an appealing question. A revival of the interest for delineation studies has been observed at the crossroads of sociology of science and analyses of networks [2.25, 26].

Disciplinary views, as well as colleges revealed by bibliometrics, lead to different partitions of literature, depending on the vantage points. In particular, bibliometricians can be confronted with conflictual situations when revealed networks and institutional normative perceptions and claims as to the disciplinary structure and boundaries diverge. The exercise of delineation generally consists in reaching some form of consensus, or at least a few consensual alternatives amongst sponsors, stakeholders, experts, and scientists. The toolbox contains information retrieval, data analysis, and mapping. Bibliometricians act as organizers of experts' supervision, suppliers of quantitative information, and facilitators of negotiations (Fig. 2.1).

2.2.2 Operationalization: Three Models of Delineation

In their review of (inter)disciplinarity issues, *Sugimoto* and *Weingart* [2.27] stress that the rich conceptualization of disciplinarity, quite elaborate in sociology and iconic of science diversity, does not imply clear operationalization solutions for defining fields. Scientists' claims and co-optation ("Mathematicians are people who make theorems" with several formulations, including a humoristic one by Alfréd Rényi), university organizations and traditions, epistemology, sociology, bibliometrics offer many entry points. The stakes associated to disciplinary interests and funding, for both scientists and policy makers may interfere with definitions. Introducing the national dimension, for example, shows that the coverage of disciplines is perceived differently in national research systems. Bibliometrics cannot capture the deep sociocognitive identity of disciplines but contributes to enlighten some of the facets that collective scientists' behavior let appear. The difficulty extends to multidisciplinarity measurement.

In practice, the description of disciplines available in scientific information systems takes the form of classification schemes at some granularity (articles, journals) from a few sources: higher education or research organizations for management and evaluation needs (international bodies or national institutions, for example

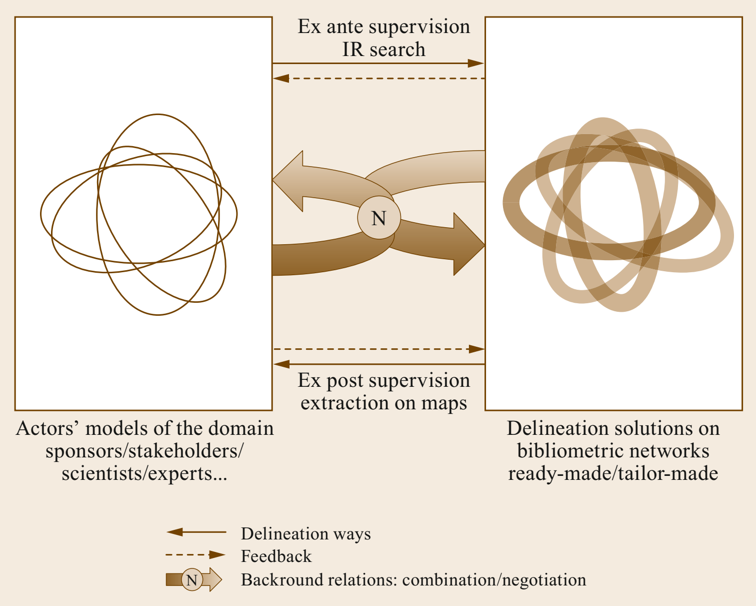


Fig. 2.1 Actors' models/bibliometric models. This scheme evokes the interaction between actors' mental or social models of science, disciplines, and domains on the one hand and models from data analyses (clustering–mapping) on bibliometric data sources, based on different methods and networks on the other. The two sides are engaged separately or together in negotiated combinations to reach (almost) consensual views. Two ways of domain delineation are singled out, ad hoc IR search and extraction from maps, with different degrees and moments of supervision. A third way, allowing direct IR search, supposes permanent classification resources

the National Center for Scientific Research (CNRS) in France); schemes associated to databases from academic societies, generally thematic; and/or from publishers or related corporations (Elsevier, ISI/Thomson Reuters/Clarivate Analytics) dedicated to scientific information retrieval.

We term *model A* the principle of these institutional science classifications, which do not chiefly proceed from bibliometrics but from the interaction between scientists and librarians. Subcategories and derived sets offer ready-made delineation solutions. The effect of methodological options, the social construction of disciplines by institutions or scientific societies, with struggles for power and games of interests are unlikely to yield convergence: the various classifications of science available, not necessarily compatible, should be taken with caution. Depending on the update system, they also tend to give a *cold* image of science. Often based on nonoverlapping schemes, they tend to handle multidisciplinarity phenomena poorly. Resources associated with classifications in S&T databases which often include various nomenclatures (species, objects) are a distinct advantage. With its limitations this model nevertheless offers a rich substance to bibliometric studies. Since the development of evaluative scientometrics in the 1970s, in the wake of Garfield and Narin's works, categories are used as bases for normalization of bibliometric measures, especially citation indicators, but classification-free alternatives exist (Sect. 2.3.2). The rigidity of classifications has an advantage, making a virtue out of a necessity, the easy measure of knowledge exchanges between categories over time. Techniques of coclassification [2.28, 29], coindex, or coword methods (see below) make it possible to transcend the rigidity of the classification scheme.

The concept of virtual college, originally thought of as micro- or mesoscale communities with informal contours, exchanging in various ways, can be generalized to communities in science networks at any scale. Since the 1980s, this is implicit in most bibliometric studies [2.30]. Global models of science, either small worlds or self-similar fractal models, are consistent with this perspective. This scheme, termed here *model C*, is the very realm of bibliometrics. Formal and institutional aspects are partly visible through bibliometric networks but need other *scientometric* information on institutional structure of science systems. Bibliometrics and also scientometrics are blind to other networks/relations such as interpersonal networks and to the complete picture of science funding and science society relations. It follows that the delineation of fields in model A, which accounts for complex mixes nontotally accessible to bibliometric networks, cannot be retrieved by *model C* approaches. The other way round, model C makes visible implicit structures ignored by the panel of actors involved in model A classification design.

For large academic disciplines, model C merely proposes high-level groupings which might emulate the categories *disciplines* from model A and share the same label, however with a quite coarse correspondence. In the practice of model C, large groups receive a sort of *discipline* label through expert supervision. Neither the bibliometric approach nor model A have the property of uniqueness. Various tests were conducted by external bibliometricians on SCI-WoS (Science Citation Index of the Web of Science) subject categories, and the agreement is not, usually, that good [2.31] and the existing ready-made classifications cannot pretend to the status of ground truth or gold standard for domain delineation. Depending on the organization, the clustering–mapping operations often fulfill two needs in bibliometric studies, first helping domain delineation, secondly identifying subdomains/topics within the target. In the absence of ground truth, the challenge of model C is to find trade-offs for reflecting a fractal reality quite difficult to break down, since boundaries are hardly natural except for configurations with clear local minima. They are then subject to optimization with partial information and negotiations [2.32, 33].

Model B based on IR search, borrows from both A and C. In model A, the operationalization of discipline definition and classification relied on heavily supervised schemes, aiming chiefly at information retrieval. Model B shares the same ground, with an ad hoc search strategy established by bibliometricians and experts for the needs of the study. Ad hoc search is sometimes necessary in order to go beyond the synthetic views provided by clustering and mapping, and to address analytical questions from users (in terms of theory, methods, objects, interpretation). The default granularity is the document level.

The three models can incorporate a semantic folder. Some indexing and classifications systems provide elaborate structures of indexes and keywords: thesaurus and ontologies (Sect. 2.2.4). Model B depends on expert's competence and resources of queried databases to coin semantically robust queries. Model C can treat metadata of controlled language, indexes of any kind, as well as natural language texts, and reciprocally shed light, through data/queries treatment, on the revealed semantic structures of universes.

Reflexivity is present under many aspects: scientists are involved in heavy ex ante input in ready-made classifications (model A), in IR ad hoc search (model B), and in softer ad hoc intervention on bibliometric maps (model C). The supervision/expertise question goes beyond within-community reflexivity, with partners associated to projects: decision-makers and stake-holders and bibliometricians.

Table 2.1 sums up the main features of the three models. They are just archetypes: in practice, blending is the rule. If classical disciplinary classification schemes belong to the first model, the Science Citation Index and variants incorporate bibliometric aspects. Purely bibliometric classifications, if maintained and widely available, give birth to ready-made solutions. In the background of the three models, the progressive rapprochement of bibliometrics and IR tools, addressed below in Sect. 2.3 should be kept in mind.

Table 2.1 Typical features of the three models for delineating scientific fields

	Model A **Ready-made breakdown and tools**	**Model B** **Ad hoc IR search**	**Model C** **Bibliometric networks**
Basic concept	Science classifications and nomenclatures	Union of queries	Groupings in science networks, generalizing the concept of invisible colleges
Origin	Academic societies and database providers. Originally little/no input of bibliometrics	Publication records (e.g., article metadata) and possibly full text	Analysis of bibliometric networks from any field in publication records
Structure	Classification schemes, often hierarchical and hard breakdown (categories: subdisciplines, fields, specialities, journals, etc.)	Categories: only when at low risk (e.g., core categories)	Networks and clusters/groups at various scales (actors, topics, documents...)
Supervision/expertise	Heavy ex ante embodied input by scientists, experts, librarians	Heavy involvement of scientists, experts, librarians in conception/check of queries	Ad hoc softer supervision at various stages (mapping)
Data–granularity	Richness of added metadata, especially keywords and indexes of objects default granularity: category	All available information, especially text fields, citation, authors' affiliations Default granularity: document	All available information, especially text fields, citation, authors' affiliations Default granularity: cluster
Semantic aspects	Thesauri, ontologies	Structure of queries, use of ready-made resources	Latent or explicit dimensions in networks
Time features	Relative stability of framework, favoring fixed-structure longitudinal analysis, at the expense of tensions in the system between updates	No structural constraint	Immediacy and aptitude to dynamic analysis of changing entities

2.2.3 Challenges at the Mesolevel

Interdisciplinarity

Interdisciplinarity is quite an old question and rose to the forefront in the early 1970s with an OECD (Organization for Economic Cooperation and Development) conference devoted to the topic, which gave rise to a wealth of literature and programs. The distinction between multi-, inter-, and transdisciplinarity formulates various degrees of integration, see [2.5, 34]. As *Choi* and *Park* put it [2.35]:

> Multidisciplinarity draws on knowledge from different disciplines but stays within their boundaries. Interdisciplinarity analyses, synthesizes and harmonizes links between disciplines into a coordinated and coherent whole.

Jahn et al. [2.36] examine two interpretations of *transdisciplinarity* in literature. Both make sense in a delineation context. One privileges the science–society relationship: integration between social sciences and humanities (SSH) and natural sciences with the participation of extrascientific actors, as a response to heavy and controversial socioscientific problems such as climatic change, genetically modified organisms, medical ethics, etc. The second interpretation considers that transdisciplinarity simply pushes the logic of interdisciplinarity towards integration. *Russell* et al. [2.37], cited by *Jahn* et al. [2.36],

> emphasize that where interdisciplinarity still relies on disciplinary borders in order to define a common object of research in areas of overlap [...] between disciplines, transdisciplinarity truly transgresses or transcends [them].

Klein [2.38] and *Miller* et al. [2.39] stress the theoretical and problem-solving capability of the transdisciplinary view. Many publications evoke the paradox of multidisciplinarity, a source of radical discoveries, laboring however to convince evaluators in the science reward system. *Yegros-Yegros* et al. [2.40] list a few controversial studies on the topic, and note a specific difficulty for distal transfers. *Solomon* et al. [2.41] recall that the impact of many multidisciplinary journals is misleading in this respect, since their individual articles are not especially multidisciplinary.

Bibliometric operationalization has to account for those different multi/inter/transdisciplinarity forms. *Multidisciplinarity* involves sustained knowledge exchanges in a roughly stable structure; *interdisciplinarity*, with an organization and systematization nuance, supposes strong exchanges creating some structural strain, between domain overlap and autonomization of merging fractions; *transdisciplinarity* paves the way for the autonomy of the overlapping region, within the strong interpretation involvement of SSH and possibly of extrascientific considerations. Clearly model C is apter than A to depict those forms and their transitions when they occur, rather than waiting for the institutionalization of the emerging structures.

Interdisciplinarity may be outlined at the individual level by copublications of scholars with different educational or publication backgrounds, by measures of knowledge flows (citations), contents proximity, authors' coactivity or thematic mobility—if such data exist [2.42]. Other sources include joint programs, joint institutions or labs claiming disciplinary affiliation, generally found in metadata. Most disciplinary databases lagged behind the Garfield SCI model as to the integral mention of all authors' affiliations on an article. The large scope of bibliometric measures of multidisciplinarity was reviewed in many articles, e.g., [2.27, 43].

In model A the first entry point to multidisciplinary phenomena is the category classification schemes, with measures of knowledge exchanges by citation flows between categories (*Pinski* and *Narin*'s seminal work on journal classifications [2.44], *Rinia* et al. [2.45]), transposable to textual proximity (on patents [2.46]) or authors coactivity. Despite the heavy input of experts in science classification, the delimitation of particular fields varies across information providers and none can be held as a gold standard. It finds its limits in the inertia and often the hard scheme of classes, albeit the derived coclassification and coindex treatments noticed above relax the constraint and instil some of the bibliometric potential of model C.

Model C is more realistic in depicting the combinatory, flexible, multinetwork relationships in science and the demography of topics. Ignoring disciplinarity as such, it conveys a broader definition of interdisciplinarity, ranging from close to distant connections, the latter loosely interpretable, in the common acceptation, as interdisciplinary and possibly forerunners of more integrated relations. More generally, the network perspective of model C builds bridges between networks formalization and scientific communities life, leaving open the question of how profoundly the sociocognitive phenomena are captured. Data analysis methods such as correspondence analysis (CA), latent semantic analysis (LSA), latent Dirichlet allocation (LDA) addressed below, claim light semantic capabilities at least. Bibliometrics cannot substitute for sociological analysis, which exploits the same tools but goes further with specific surveys. Similarly, it is dependent on computational linguistics and semantic analysis for deep investigations of the knowledge contents. Model C is a potential competitor for offering taxonomies, with recent advances (Sect. 2.2.4). It does not follow that dynamics captured by this model are easy to handle: for example, flow variations in a fixed structure (A) read more conveniently than multifaceted structural change (C).

Internal Diversity

Diversity and multidisciplinarity are two facets of a coin. Internal diversity in a delineation process qualifies communities inside the target domain. Figure 2.2b,c expresses the internal diversity of multidisciplinary domains, already striking for nanosciences and massive for proteomics (Fig. 2.2).

Internal diversity is treated in quite different ways depending on the model. In the cluster analysis part of model C, the balance of internal diversity and external connectivity (*multidisciplinarity* in the looser sense) is part of the mechanism which directly or indirectly rules the formation of groups, with a wide choice of protocols. Many solutions of density measurement are available in clustering or network analysis, with some connections with diversity measures developed in ecology and economics especially. The synthetic Rao index discussed by *Stirling* [2.49] combines three measures on forms/categories: variety (number of categories), balance (equality of category populations), and disparity (distance of categories). Delineation through mapping will use smaller scale clusters rather than attempting to capture the target as a whole large-scale cluster. There is no risk of missing large parts of the domain, but the way the different methods conduct the process raises questions about the homogeneity of clusters obtained and the loss of weak signals especially in hard clustering (Sect. 2.3).

In model B internal diversity, especially when generated by projected multidisciplinarity, is a threat on recall. Entire subareas may be missed out if the diversity in supervision (panels of experts) does not match the diversity of the domain. Unseen parts will alter the results. In contrast, on prerecognized areas, model B can be tuned to recover weak signals.

In model A, the existence of a systemic silence risk particularly depends on how interdisciplinary bridges are managed.

Unsettlement

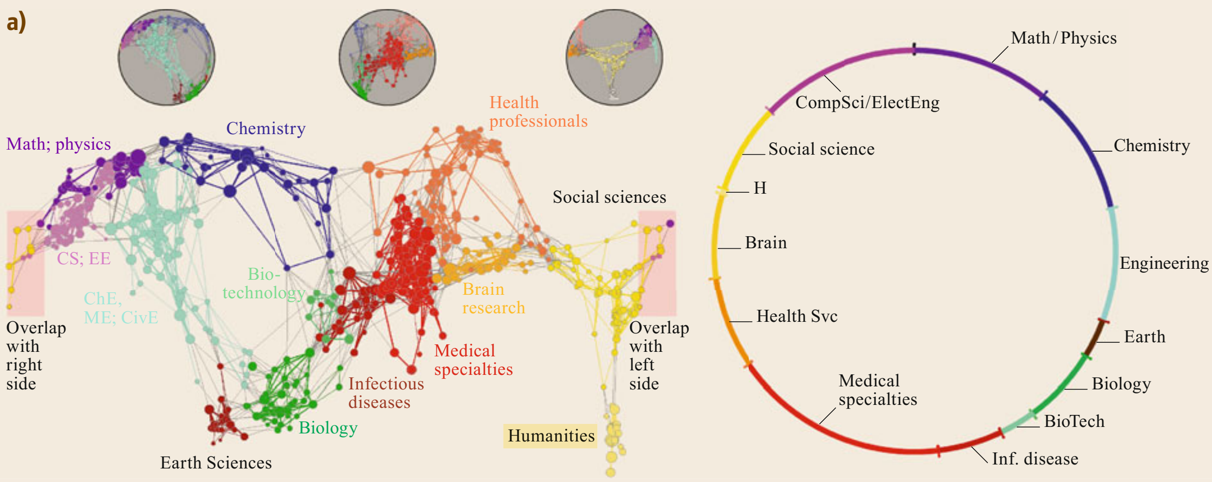
The third challenge of domain delineation lies in the science network dynamics. Conventional model A classifications hardly follow evolutions and need periodic adjustments. The convenience of measures within a fixed structure is paid for by structural biases. Bibliometric mapping can translate evolutions in cluster or factor reconfiguration, but the handling of changes in a robust way remains delicate (Sect. 2.3). Model B pictures networks, but intuitively, a fast rhythm of reconfiguration in the somewhat chaotic universe of science networks makes it particularly difficult to settle delineation on firm roots. This casts a shadow on the time robustness of the solutions reached on one-shot exercises, but also on the predictive value of extrapolations on longitudinal trends. We will return later to dynamic studies and semantic characterization (Sect. 2.3.2). Emerging domains seldom embody institutional organization but bear bibliometric signatures of early activity. The difficulty is to capture weak signals with a reasonable immediacy. Fast manifestations of preferential attachment around novel publications, whatever the measure (citations, concept markers, or altmetric linkages) are amongst the classical alerts of topic emergence at small scale, to confirm by later local cluster growth.

Source Coverage

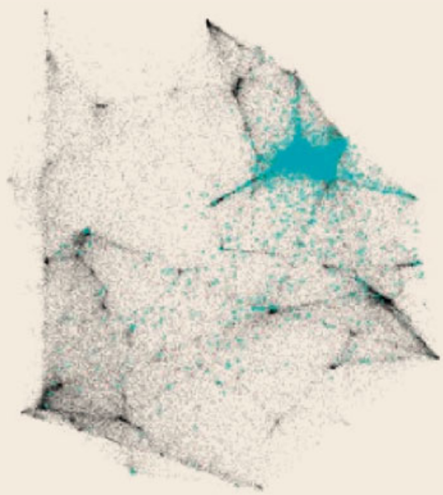
For memory's sake, the question of data coverage is recurrent in practical bibliometrics and is raised at the delineation stage of any study. The literature on the subject is abundant, conveying different points of view: *Hicks* [2.50] first stressed the limitations of both the reference database SCI and the mapping algorithm of cocitation for research policy purposes. *Moed's* review [2.51, esp. Sect. 6.2.2] and *Van Raan* et al. [2.52] showed the differential coverage of disciplines by journals in SCI-WoS using references to nonsource items. Keeping pace with the growth of visible science is another challenge. The latest United Nations Educational, Scientific and Cultural Organization (UNESCO) science report estimates that 7.8 million scientists worldwide publish 1.3 million publications a year [2.53]. SCI-WoS producers proposed new products beginning to fill the gap of book literature, essential to social sciences and humanities (SSH) and conference proceedings, essential to computer science [2.54]. The coverage of social science and humanities with issues of publication practices and national biases was addressed in many works, e.g., [2.55–57]. This is distinct from the within-discipline approach where an extensive coverage causes instability of indicators due to tails (language biases, national journals biases), to document types or adaptation issues [2.58–61]. Former studies' figures are outdated but the basic principles remain.

Extensive databases with enhanced coverage for IR purposes (modern WoS, Scopus) might require truncation of tails for comparative international studies. The PageRank selection tool limits the noise of a massive extension of sources in Google. However, Google Scholar is not considered a substitute for biblio-

a)

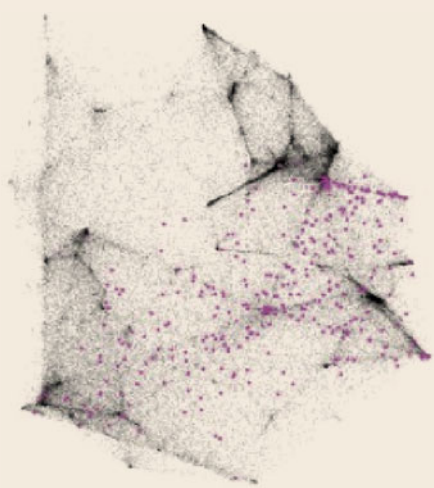


b)



Nanotechnology
Most research communities in nanotechnology are concentrated in **physics**, **chemistry**, and **materials science**, However, many disciplines in the life and medical sciences also have nanotechnology applications.

c)



Proteomics
Research communities in proteomics are centered in **biochemistry**. In addition, there is a heavy focus in the tools section of chemistry, such as **chromatography**. The balance of the proteomics communities are widely dispersed among the life and medical sciences.

Fig. 2.2a–c Map of science and multidisciplinary projections. **(a)** A world-map-type science map from a spherical representation [2.47]. **(b)** and **(c)** Hotspots of activity of nanoscience and proteomics projected in a fraction of a global science map. It basically crosses the map's holistic picture with an overlay of hits from simple term queries. After *Boyack* and *Klavans* [2.48] ◄

graphic databases for common librarian tasks, but rather a complement especially for coverage extension in long tails [2.62] with variations amongst disciplines. The same applies to another large bibliographic database: the Microsoft Academic Graph [2.63–65]. The lack of transparency in the inclusion process and the lack of tools beyond original ranking (sorting, subject filters) are stressed by *Gray* et al. [2.66]. Strong concerns with the quality of bibliographic records were also reported [2.67, 68]. The coverage of databases has recently been compared by several authors [2.69, 70], with an extension to alternative sources such as altmetrics: http://mendeley.com, http://academia.edu, http://citeulike.org, http://researchgate.com, http://wikipedia.org, http://twitter.com, etc. [2.71, 72]. Online personal libraries like Mendeley shed new light on knowledge flows between disciplines through publication records stored together [2.73]—a kind of cocitation data from readers instead of authors. In addition, these sources, often difficult to qualify properly [2.74], have been addressed by altmetric studies [2.42, 75, 76]. The way scientists and the general public communicate about science on (social) media is field-dependent and it is not easy for now to anticipate the complementary role of altmetrics and traditional data in delineation of fields. Altmetric resources can help exploratory and supervision tasks.

In emerging and multidisciplinarity topics that typically justify careful delineation, controversies and conflicting interests are frequent and the importance of transdisciplinary problems makes the issues of sources coverage, experts panel selection, and supervision organization more acute.

2.2.4 Ready-Made Classifications

Classifications

Table 2.2 presents some types of science classifications valuable in domain delineation. These coexisting classification schemes reflect various perspectives, such as cognitive, administrative, organizational, and qualification-based rationales according to *Daraio* and *Glänzel* [2.77] who stress the difficulties arising when trying to harmonize them.

The first named classifications directly stem from professional expertise of scientists and librarians (pure model A). Some are linked to institutional or national research systems, mainly oriented towards staff management or evaluation, or international instances (UNESCO). More relevant for bibliometric uses are classifications within complete information systems on S&T literature, proceeding from a few sources: specialized academic societies (CAS (Chemical Abstracts Service), Inspec, Biosis, MathSciNet, Econlit, etc. which usually extend beyond their core discipline) and/or scientific publishers, and patent offices for technology. Classifications are typically hierarchical, complemented by metadata (keywords of various kinds, indexes from object nomenclatures: vegetable or chemical species, stellar objects, and so on).

Table 2.2 Science classifications

1.	International classifications, often high level. OECD high-level a.k.a. *Frascati Manual*. Fields of science introduced in 2002. Last revision in 2007. Correspondence table with WoS [2.78]. Six major fields were subcategorized.
2.	Institutional nomenclature frameworks (ex. CNRS sections[a]). Reflects the vision of the institution and its involvements.
3.	Bibliographic databases from science societies. Involve nomenclatures and/or classifications, with a disciplinary focus, sometimes very large (ex. *Chemical Abstracts Service* CAS). Typically based on a classical documentation system, with heavy expert input. Another example of classification in computer science: *association of computer machinery classification*[b] 1964–2012.
4.	Alternative *ISI model* ISI (Institute for Scientific Information)/Thomson/Thomson Reuters/Clarivate; Scopus/SCImago Journal Rankings[c] (SJR) as "a publicly available portal that includes the journal and country indicators developed from the information contained in the Scopus database." First used the editorial entity *journal* as the basic molecule, and impact as a principle of selection (see historical account by *Garfield* [2.79]). Extensions at a more detailed level. The balance expertise/bibliometrics to design subject categories is unclear (see WoS notices on the topic and [2.80, p. 1113]). Gives a one- or multilevel hierarchy of groups. The database offers both nonoverlapping schemes (essential indicators) and overlapping schemes (SCI-WoS).
5.	Bibliometric mapping classifications, either at the journal or the document level: tailor-made maps potentially usable as permanent resources for public purposes.

[a] http://www.cnrs.fr/comitenational/english/section_acc.htm
[b] http://www.acm.org/about/class
[c] http://www.scimagojr.com/aboutus.php

Bibliometrics then entered the competition for science classifications, in contrast with the traditional documentation model involving heavy manpower for indexing individual documents. The prototype is *Garfield's* SCI/WoS based on the *journal* molecule and a selection tool, the impact factor [2.81, 82]. The supervision was still heavy in the elaboration of classification, although the journal citation report is a powerful auxiliary for actual bibliometric classification based on journals' citation exchanges [2.83]. The model of citation index inspired Elsevier's Scopus [2.84, 85]. The Google Scholar alternative, with a larger scope of less normalized sources, is the extreme case with very little supervision and does not include a classification scheme.

Following Narin's works, several journal classifications were developed (factor analysis in [2.86], core–periphery clustering in [2.87]). Many others have been proposed over the past decades, some with overlay facilities for positioning activities [2.88]. Other proposals use prior categories and expert judgments as seeds [2.89, 90], with reassignment of individual papers. *Boyack* and *Klavans*, whose experience covers mapping and clustering at several granularity levels (journals, papers) [2.91], recently reviewed seven journal-level classifications (Elsevier/Scopus ASJC (All Science Journal Classification), UCSD (University of California San Diego), Science-Metrix, ARC (Australian Research Council), ECOOM (The Center for Research and Development Monitoring), WoS (Web of Science), NSF, JID (Journal IDentification) and ten article-level classification (five from ISI and Center for Research Planning (CRP), four from MapOfScience, one from CWTS (Center for Science and Technology Studies)) [2.92]. The latter authors privilege the concentration of references in review articles (>100 references) considered as *gold standard* literature, as an accuracy measure (a heavy hypothesis). They conclude in favor of paper-level (versus journal-level approaches) and in favor of direct citations (versus cocitations or bibliographic coupling) for long-term smoothed taxonomies, distinguished from current literature analysis, for which they rank first bibliographic coupling.

Those developments mark a new turn in the competition between institutional classification and bibliometric approaches for long-term classifications of science. It is not clear, however, whether the variety of classifications from bibliometric research, not always publicly available, can supersede the quasistandards of SCI type for current use in bibliometric studies. High-quality delineation of fields cannot solely rely on journal-level granularity, and this is still more conspicuous for emerging and complex domains.

Semantic Resources

Science institutions and database producers have a continuous tradition of maintenance of linguistic and semantic resources, in relation to document indexing. The best known is probably the MeSH (Medical Subject Headings; National Library of Medicine) used in Medline/PubMed. INSPEC, CAS, and now Public Library of Science (PLoS) offer such resources. Controlled vocabulary and indexes, archetypal tools of traditional IR search were also the main support of new coword analysis in the 1980s. A revival of controlled vocabulary and linguistic resources is observed in recent works, associated to the description of scholarly documents [2.93] and bibliometric mapping [2.94]. We shall return to the role of statistical tools in the shaping of semantic resources.

2.2.5 Conclusion

Science, seen through scientific networks, is highly connected, including long-range links reflecting interdisciplinary relations of many kinds. Global maps of science, with the usual reservation on methods settings and artifacts, display a kind of continuity of clouds along preferential directions (Fig. 2.2c, from [2.47]). The extension of domains has to be pragmatically limited by IR trade-off with the help, in the absence of ground truth, of more or less heavy supervision. Three models of delineation appear: ready-made delimitation in databases, rather limited and rigid as is, but prone to creative diversions from strict model A (coclassification, etc.); model B, ad hoc search strategies combining several types of information; model C, by extraction of the field from a more extended map, regional or global.

Networks of science may locally show cases of domains ideal for trivial delineation: a perfect correspondence between the target and ready-made categories, or insulated continents surrounded by sea. Such domains will not require sophisticated delineation. This is the exception not the rule.

Areas such as environmental studies, nanosciences, biomedicine, information and cognitive sciences and technologies (converging NBIC, concept coined by NSF in 2002) exhibit both internal diversity and strong multidisciplinary connections. Commissioned studies often target emerging and/or high-tech strategic domains which witness *science in action* prone to socio-scientific controversies à la Latour. These areas combine high levels of instability and interdisciplinarity. As to transdisciplinarity, the question arises of whether to include SSH and alternative sources in data sources and panels of experts.

2.3 Tools: Information Retrieval (IR) and Bibliometrics

This section focuses on some technical approaches to the delineation problem: information retrieval and bibliometric mapping. They share the same basic objects and networks, chiefly actors and affiliations, publication supports, textual elements, and citation relations. Although the general principles of bibliometric relation studies are quite well established, new techniques from data analysis and network analysis, including fast graph clustering, open new avenues for achieving delineation tasks on big data at the fine-grained level. The quality of results remains an open issue. Domain delineation confronts or combines the three approaches previously stated: ready-made categories (model A) are seldom sufficient; we shall envision ad hoc IR search (model B) with an occasional complement of ready-made categories; and bibliometric processes of mapping/clustering along the lines of model C.

2.3.1 IR Term Search

The question of delineation spontaneously calls for a response in terms of information-retrieval search. The only particularity is the scale of the search or more exactly, as mentioned before, the diversity expected in large domains, which is particularly demanding for the a priori framework of information search. The verbal description of the domain requires, beforehand, an intellectual model of the area. In addition to the methodological background brought by IR models, a broad range of search techniques address delineation issues:

- Ready-made solutions in the most favorable cases, with previously embodied expertise, sketched above.
- Search strategies of various levels of complexity, also depending on the type of data, relying on expert's sayings.
- Multistep protocols: Query expansion, combination with bibliometric mapping.

IR models are outside the scope of this chapter. In the tools section below, we recall some of the techniques shared by IR and bibliometrics, especially the vector-space-derived models.

IR Tradeoff at the Mesolevel

The recall–precision trade-off is particularly difficult to reach at the mesolevel of domains exhibiting high diversity. Generic terms (say the *nano* prefix if we wish to target nanosciences and technology) present an obvious risk to precision. A collection of narrower queries (such as *self-assembly*, *quantum dots*, etc.) is expected to achieve much better precision. In the simpler Boolean model, this will privilege the union operator of subarea descriptors (examples for nanoscience [2.33, 95, 96]). However, nothing guarantees a goodness of coverage of the whole area by this bottom-up process. An a priori supervision of the process by a panel of experts is required, but the experts' specialization bias, especially in diverse and controversial areas, generates a risk of silence. Similar risks are met in the selection of training sets in learning processes. Another shortcoming is the time-consuming nature of supervision, again worsened by the diversity and multidisciplinarity of the domain. A light mapping stage beforehand may reduce the risk of missing subareas. As mentioned above, focused IR searches are, in contrast, able to retrieve weak signals lost in hard clustering.

Polyrepresentation and Pragmatism

Scientific texts contain rich information, most of it made searchable in the digital era. Pragmatically, all searchable parts of a bibliographic record, data or metadata are candidates for delineating domains: word *n*-grams in titles, abstracts, and full texts; authors, affiliations, date, journal or book, citations, acknowledgements, transformed data (classification codes, index, controlled vocabulary, related papers. . .) depending on the database. These various elements exhibit quite different properties. In theoretical terms, the variety of networks associated to these elements are one aspect of the *polyrepresentation* of scientific literature [2.97]. We will return to this question later (Sect. 2.3.2). A specific advantage of lexical search is the easy understanding of queries—whereas other elements (aggregated elements such as journals; citations) are more indirect. However, the ambiguity of natural language reduces this advantage.

Bibliometric literature is packed with examples of pragmatic delineation of domains based on IR search. By and large, apart from ready-made schemes when available (indexes, classification codes), a typical exploration combines a search for specialized journals if any, and a lexical search in complement. At times, an author-affiliation entry is used, especially in connection with citation data. Bradford and Lotka ranked lists are therefore good auxiliaries, with evident precautions on journals' or authors' degree of specialization.

Granularity

We noted above that some ready-made classifications such as the SCI scheme (journals or journal issues) are essentially based on full journals—or journal sections. These ready-made categories very seldom fit the needs of targeted studies. Instead, ad hoc groupings of selected journals relatively easy to set up with the help of experts, are a convenient starting point within a Bradfordian logic. The journal level presents obvious advantages. Journals exhibit a relative stability in the medium term; they are institutionalized centers of power through gatekeeping, and a (controversial) evaluation entity in the impact factor tradition.

However, the journal level is problematic for delineation studies. Journals whose specialization is such that they indisputably belong to the target domain, can be taken as a whole, but of course target domain literature are rarely covered by specialized journals only, and investigations should be extended to moderately or heavily multidisciplinary sources. Conditions of diversity and multidisciplinarity—which prevail in the targets of studies where elaborate delineation is worthwhile—hinders the efficiency of global Bradford/Lotka-based selections, with problems of normalization (refer also to [2.98]). We will return to these issues in the Sect. 2.3.2 devoted to clustering and mapping.

To conclude, the IR resources in scientific texts, data and metadata, suggest a polyrepresentation of scientific information (cognitive model [2.97]), which is akin to the multinetwork representation of the scientific universe. *Ingwersen* and *Järvelin* [2.99, p. 19] propose a typology of IR models and the perspective of the *cognitive actor*. IR protocols generally involve multistep approaches, with various core–periphery schemes. In conventional search, heavy ex ante supervision is needed for covering the variety of domains, ideally with good analytic/semantic capability. In the absence of a gold standard, proxy measures of relevance are needed.

Multistep Process

Multistep processes, possibly associated with combinations of various bibliometric attributes, are run-of-the-mill procedures (for example [2.32]).

Core–periphery rationale is common, in accordance with the selective power of concentration laws, both in IR and bibliometrics (journal cores in [2.100], cocitation cores in [2.101], *h*-core in [2.102], emerging topics in [2.103]). For example, working on highly cited objects—authors, journals, or articles—gives a set of reasonable size, amenable to further expansion with enhanced recall. Cores inspired from the Price law on Lotka distributions or from application of the *h*-index are helpful. Proxies such as seeds obtained from initial high-precision search stages can do as well. The core or seed expansion process is global or cluster-based. The risk of core–periphery schemes, by and large favorable to robustness, is to miss lateral or emerging signals. This may need some input of dynamic characterization of hotspots at the fine granularity level.

A parent method is bibliometric expansion on citations, which also uses information from a first run (set of documents retrieved by a search formula or a prior top cited selection, considered as the core) to enhance the recall through the citation connections, typically operating at the document level with or without a clustering/mapping step. In this line the Lex+Cite approach mentioned in Sect. 2.3 relies on a default global expansion, rather than a cluster-based one, to limit the risk of an exclusive focus on cluster-level signals that would miss across-network bridges.

Query expansion by adaptive search is along the same lines. Interactive retrieval with relevance feedback identifies the terms, isolated or associated (cooccurrences), specifically present in the most relevant documents retrieved according to various measures [2.104–106]. An efficient but heavy process consists in submitting the output of a search stage to data analysis/topic modeling, able to reconstruct the probable structure likely to have generated the data. By providing information on the linguistic context—also citation, authoring context, etc.—they in turn help to improve the search formulas by a kind of retro-querying. This ranges from simple synonym detection to construction of topics, orthogonal or not, suggesting the rephrasing of queries. Variants of itemset mining uncovering association rules ([2.107], with earlier forerunners) are promising in this respect (see below). Evaluation of output from unsupervised stages can also call for a manual improvement of queries.

Delineation protocols may also use the seed as a training set for learning algorithms. A difference is that core–periphery schemes usually rely on the selective power of bibliometric laws, whereas the training set might be extracted on various sampling methods, provided that the seed does not miss the variety of the target. As big data grows bigger, *semisupervised* approaches are gaining popularity in the machine learning community. This recent approach should prove attractive in the bibliometrics community, as there seems to be considerable interest in linking metadata groups and algorithmically defined communities [2.108].

To conclude on this part, whilst typical IR search relies on an a priori understanding of the field, multistep schemes involve stages of data analyses quite close to bibliometric mapping practices, the topic of the next subsection. IR and bibliometrics share roots and features, which soften the differences: adaptive loops, learning processes, seed-expansion, and core–periphery schemes. Bibliographic coupling, at the very origin of bibliometric mapping, came from the IR community [2.109] and the *cluster hypothesis* about relevant versus nonrelevant documents [2.110] voices the common interests of IR and bibliometrics, beyond the background methodology of information models (Boolean, vector space, or probabilistic) and general frameworks such as the above-mentioned cognitive model. The tightening of bibliometrics–IR relations has been echoed in a series of workshops and in dedicated issues of *Scientometrics* ([2.111, 112], see also [2.113] for a focus on domain delineation) and in the *International Journal on Digital Libraries* [2.114].

2.3.2 Clustering and Mapping

In contrast with conventional IR search, bibliometric mapping starts at a larger extension level than the targeted domain. This broad landscape, typically built by unsupervised methods, is scrutinized by experts to rule out irrelevant areas. The supervision task is limited to the *postmapping* stage. This is in principle less demanding than the a priori conception of a search formulation or of a training set. The default solution is a zoomable general or regional map of science, with availability and cost constraints. The alternative is the construction of a limited overset including almost certainly the anticipated domain, using a general search set for massive recall, an operation much lighter than setting up a precise search formula. In terms of scale, the final result is tantamount to the outcome of a top-down elimination process, although the selection modalities are diverse. There is currently great interest in delineation through mapping. IR and mapping are complementary in various ways. Firstly, we briefly describe the data analysis toolbox, before addressing the main bibliometric applications and a few problematic points.

Background Toolbox

The data structure of matrices in the standard bibliometric model allows scholars to mobilize the large scope of automatic clustering, factor/postfactor methods, and graph analysis. Classical methods of clustering and factor analysis continue to be used in bibliometrics, but in the last decade(s) novel methods came of age, more computer-efficient and fit for big data, an advantage for mapping science and delineating large domains. Starting with bibliometric data of the standard model and some metrics of proximity or distances, clustering and community detection methods produce groups. Elements are mapped using various dimension reduction algorithms. Factor methods produce groups through clustering applied to factor loadings, with an integrated two-dimensional (2-D) or three-dimensional (3-D) display when just two or three factors are needed in the analysis.

A major driving force of bibliometric methodology is the general network theory, which took large networks of science, especially collaboration and citation, as iconic objects [2.115–117]. Quite a few mechanisms have been proposed to explain or generate scale-free networks since *Price's* cumulative advantage model for citations [2.118] along the lines of Yule and Simon, and later studied in new terms (preferential attachment) by *Albert* and *Barabási* [2.119], see also [2.120]. These models have some common features with the Watts–Strogatz small worlds model, but also differences that are empirically testable [2.121]. Amongst other mechanisms: homophily [2.122], geographic proximity [2.123], thematic proximity inferred from linguistic or citation proximity. *Börner* et al. reviewed a few issues in science dynamics modeling [2.124]. Of great interest in bibliometrics and especially delineation, community detection algorithms exhibit a general validity beyond *real* social networks, and belong to the general toolbox of mathematical clustering and graph theory—applicable to various markers of scientific activity, document citations, words, altmetric networks, etc., see also [2.120].

Hundreds of clustering and mapping methods have been designed during one century of uninterrupted research. This section can only provide a basic overview of the main method families, in the perspective of domain delineation. More comprehensive descriptions and references, as well as a basic benchmark of various methods, applied to a sample of textual data, can be found in [2.125].

Clustering Methods. Although *hierarchical clustering* algorithms sometimes seem old-fashioned because of their computing complexity, $O(n^2)$ in the very best cases, some of them show good performances for relative small universes. For large ones, they can be coupled beforehand to data-reduction stages, classical (SAS Fastclus $O(n)$), preclustering algorithms for big data (Canopy clustering [2.126]), or sampling methods. All-science bibliometric maps use rather faster algorithms today, not without limitations however. Discipline-level maps, or simply internal clustering of the domain set at various stages of delineation may still rely on the classical techniques.

Hierarchical ascending algorithms are local, deterministic and produce hard clusters, with a few exceptions (pyramidal classification), properties favorable to dynamic representations. They do not constrain the number of clusters and provide a multiscale view through embedded partitions, with some indication of robustness of forms in scale changes. Most hierarchical descending (divisive) methods are heavier. Hierarchical methods typically rely on ultrametrics, which has downsides, see [2.125].

Amongst popular methods in bibliometrics are ascending methods: single linkage, average linkage, and

Ward. Single linkage is relatively fast and exhibits good mathematical properties in relation to spanning trees but produces disastrous chain effects which must be limited in various ways. Ward and especially group average linkage give better results. Group average linkage, advocated for bibliometric sets by *Zitt* and *Bassecoulard* [2.127] and used by *Boyack* and *Klavans* in various works [2.128], is slightly biased towards equal variance and is not too sensitive to outliers. Ward is biased towards equal size with a strong sensitivity to outliers. Properties and biases were studied especially by *Milligan* [2.129, 130] using Monte Carlo techniques.

Density methods are appealing: deterministic too, local, and as such prone to dynamic representations of publication or citation flows. DBSCAN [2.131] (density-based spatial clustering of applications with noise) is the most popular to the point of becoming synonymous with *density clustering*. The SAS clustering toolbox includes hierarchical methods with prior density estimation, with good properties towards sampling and the ability to capture elongated or irregular classes. However, this property is disputable in bibliometric uses (Sect. 2.3.2, *Shape/Properties of Clusters*). More recently, density peaks [2.132] has implemented an original and graphical semiautomatic procedure for determining the cluster seeds.

Not directly hierarchical is the venerable *K-means clustering* family, still popular, thanks both to its excellent time/memory performance and sensitivity to different cluster densities. A shortcoming of not being deterministic, they converge to local optima of their objective function, depending on their random (or supervised) initialization. In comparative analyses, they are not considered too sensitive to outliers. They optionally allow for soft/fuzzy clusters, and approximate dynamic data-flow analysis.

Factor methods are basically dimension-reduction techniques, indirectly linked to the partition problem. A quick-and-dirty heuristics for extracting a limited number k of dominant clusters from k factors consists of assigning each entity to the factor axis which maximizes the mode of its projection, subject to the constraint of a common factor sign for the majority of entities assigned to this cluster—which eliminates few of them in practice. For a more rigorous procedure, see the descending hierarchical clustering method Alceste [2.133] in the dataspace of correspondence analysis. Factor methods rely on the mathematical foundation of singular value decomposition (SVD) of data matrices for reducing dimensionality and filtering noise. The interesting metrics used by correspondence analysis (CA [2.134]) explains the attention over half a century from many scholars in relation to mapping or clustering limited to a few dominant factor dimensions. Dropping this limit, i. e., taking into account factor spaces with hundreds of dimensions [2.135], latent semantic analysis (LSA [2.136]) unleashed the potential of singular value decomposition and fostered the integration of semantics in textual applications, in a lighter but more convenient form than handmade ontologies, costly to edit and update.

Hybrid factor/clustering methods, sometimes coined *topic models*, result in representing each cluster as a local, oblique factor, with a progressive scale from core elements to peripheral ones, opened to fuzzy or overlapping interpretations or extensions. Generally powered by the expectation maximization algorithm (EM), they converge to local optima, too. Non-negative matrix factorization (NMF) and self-organizing maps (SOM) are well-known examples. Axial k-means (AKM in [2.137]) has been used in a comparative citations/words bibliometric context (Sect. 2.4).

Also known as topic models, the *probabilistic models* try to lay solid statistical foundations for their hybrid-looking representation: they produce explicit generative probabilistic models for the utterance of topics and terms [2.138]. Probabilistic LSA (pLSA in [2.139]) and latent Dirichlet allocation (LDA in [2.140]) are the best-known examples, claiming good semantic capabilities. The older fuzzy C-means method (FCM) is akin to this family, which uses the EM scheme for converging to local optima of their objective function.

The *graph clustering family*, also known as *network analysis*, or *community detection methods*, does not operate on the raw (entities $\times$ descriptors) matrix, as the previous families do, but on the square (entities $\times$ entities) similarity matrix, whose visual counterpart is a graph. Most of these methods operate directly on the graph, detecting cliques or relaxed cliques (modal classification), e.g., Louvain [2.141], InfoMap [2.142], and smart local moving algorithm (SLMA in [2.143]). Some of them operate on the reduced Laplacian space drawn from the graph (spectral clustering [2.144]). Quite a few comparative studies are available [2.145–147].

Note on Deep Neural Networks. While neural networks were somewhat in standby mode during the 1995–2005 decade, challenged by more manageable mathematical methods, several factors like the pressure of big data availability and progress in hardware (GPU, i. e., graphics processing units) triggered a renewal under the banners *deep neural nets* and *deep learning*. Allowing learning by backpropagation of errors in many layers networks, they gave form to the dream of knowledge acquisition by growing levels of abstraction: for images, extraction of local features; contours, homogeneous areas, shapes; for written language: character n-grams, words, word n-grams, expressions/phrases, sentences. Typically, they avoid heavy natural language processing (NLP) preprocessing (parsing, unification, weighting, selection...). These techniques are already widely used in supervised learning, with spectacular progress in automatic translation, face recognition, listening/oral comprehension, with important investment from the largest internet-related companies (e.g., Google, Apple, Facebook, Amazon), especially. As far as informetrics and IR are concerned, the main domain impacted so far is logically large-scale retrieval ([2.148] which uses a robust letter-trigram-based word-n-gram representation). There have also been some attempts in relation to nonsupervised processes for information retrieval [2.149].

A promising technique is neural word embeddings (NWE). Millions of texts now available online make it possible to develop vector representations of words in a semantic space in a more elaborate way than LSA—a method coined *neural word embeddings*. For example, the Word2Vec algorithm [2.150] processes raw texts so as to list billions of words-in-context occurrences (e.g., word + previous word + next word), then factorize [2.151] the word × context matrix (tens of thousands of words, a few hundreds of thousands, or millions of unique contexts) and extract some hundreds or thousands of semantic and syntactic dimensions. We will return later to the semantic capabilities of NWE.

Note on the Definition of Distances. Whether starting from a binary presence/absence matrix or from occurrence or co-occurrence counts, some methods embed a specific weighting scheme, i. e., a metric, for computing distances, or similarities between items. This is the case of probabilistic models, correspondence analysis, and axial K-means. Other methods allow for a limited and controlled choice, as aggregative hierarchical methods do. In the case of graph clustering methods, the user may freely choose his preferential distance definition prior to building the adjacency matrix, which adds an extra degree of freedom beyond the choice of the degree of nonlinearity, via a threshold value. For word-based matrices, heavier than citation-based ones, the methods of the k-means family also make it possible to choose a weighting scheme (Salton's term frequency-inverse document frequency (TF-IDF), Okapi BestMatch25 [2.152]).

Whereas factor/SVD methods combine the metrics and mapping capability, e.g., two-factor planes or 3-D displays, at the native granularity level (e.g., document × words), other mapping algorithms may operate on rectangular or on square (distance) matrices of elements or on groups from a clustering stage, or institutional aggregates (journals). Families of mapping techniques rely on various principles: equilibrium between antagonistic forces—repulsion between nodes, attraction alongside edges (e.g., the Fruchterman and Reingold algorithm [2.153], implemented in Gephi [2.154], alone or combined with clustering (Sandia VxOrd/DrL/OpenOrd [2.155], CWTS VOSviewer [2.143])); optimization of diverse functions: projection stress minimization in the case of MDS (multidimensional scaling), with Euclidean distances in the case of metric MDS, a variant of PCA (principal component analysis), and other distances or nonlinear functions of these distances in the case of nonmetric MDS, one of the nonlinear unfolding techniques; maximizing inertia in the case of Correspondence analysis, minimizing edge-cuts in a 2-D projection plane; or maximizing local edge densities [2.156].

Itemset Techniques. Itemset techniques are used for describing a data universe in terms of simple procedures, typically Boolean queries with AND, OR, and NOT operators. This may be used for building a stable procedural equivalent of data, e.g., for updating a delineation task (like probabilistic factor analyses). It may also be used for query expansion, as mentioned above in Sect. 2.3.1. The problem amounts to duplicating a reference partition in a new universe: machine learning techniques are basically fit to this problem, and, in the particular context of textual descriptions, itemset techniques. They are akin to generating Boolean queries with AND, OR, and NOT operators, for extracting approximations of the delineated domain, within precision and recall limits established in the machine learning phase [2.107, 157].

A Benchmark. To illustrate the capabilities of these various methods with an example, in the absence of a bibliometric dataset labeled with indisputable *ground truth* classes, we turned towards a reference dataset popular in the machine learning community, the Reuters 21 578 ModApté split (the corpus description is available online at http://www.daviddlewis.com/resources/testcollections/rcv1/. The website http://www.cad.zju.edu.cn/home/dengcai/Data/TextData.html has made a preprocessed version of this corpus available to the public, as supplementary material to [2.158]). The main features are:

- Source: A set of short texts: newswires from Reuters' press.
- Contents: In the six-class selection used, the number of texts (≈ 7000) and terms (≈ 4000) is sufficient with regards to text statistics.
- Class structure considered as ground truth: Built by experts, visually glaring in Fig. 2.3: two big classes,

one very dense, the other not, and four small classes, two of which are linked together. In this way, two major problems of real-life datasets are addressed: the imbalance between cluster sizes, and between cluster densities.

We challenge 17 clustering/mapping methods to retrieve this class structure. The similarity of their cluster solution to ground truth partition is measured by two indicators, adjusted Rand index (ARI [2.159]) and normalized mutual information (NMI [2.160]). The results are detailed in [2.125]. Let us summarize them in a user-oriented view, sorted by number of required parameters: the lesser the better, ideally, facing a bibliometric dataset without prior knowledge, no parameter:

- Two methods of network analysis require no internal parameterization, Louvain and InfoMap. However, the similarity matrix generally requires a threshold setting, here fixed to 0.1 in the cosine intertext similarity matrix. Infomap obtains the best result in terms of NMI (0.436 value versus 0.423), the index considered the best match for human comparison criteria. This value is rather poor, and this method does not distinguish classes 1, 2, 3, 4, and splits class 6.
- Nine methods require one parameter: The three hierarchical clusterings need a level cut parameter, possibly adjusted for 6 resulting clusters, while for CA, NMF, AKM, pLSA (probabilistic latent semantic analysis), LDA and spectral clustering, the number of desired clusters (6) has to be specified. As the latter group converges to local optima, we kept the best results in terms of their own objective function out of 20 runs. The indisputable winner is average link clustering, in both ARI (0.62) and NMI (0.71) terms. The lists of the four following challengers are contrasted: with regard to ARI, first Mac Quitty hierarchical clustering (0.50), then LDA, AKM, CA; with regard to NMI, first AKM (0.51), then Mac Quitty, CA, LDA. If one optimizes ARI over all 20 runs with prior knowledge of the six-clusters structure—a heroic hypothesis—, average link clustering still performs best (with a ten-clusters cut, ARI = 0.71, NMI = 0.64) while the followers reach, at best, ARI = 0.55 and NMI = 0.55.
- The last group of methods (ICA (independent component analysis), DBSCAN, FCM, affinity propagation, SLMA, density peaks) require at least two parameters, a handicap in the absence of prior knowledge of the corpus structure. SLMA obtains the best rating (ARI = 0.60, NMI = 0.55).

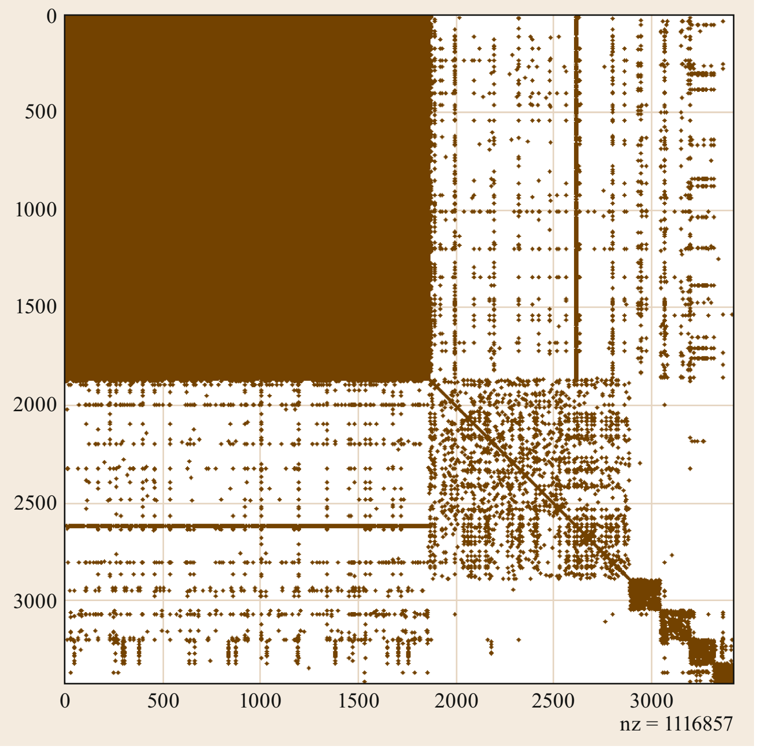


Fig. 2.3 Benchmark structure (ground truth). *Spy* plot of the cosines between document vectors of the top six classes Reuters ModApté split collection. The rows and column ordering is that of the six Reuters classes. *Black pixels* mean: cosine > 0.5

Our general conclusion is that one must be very cautious regarding domain delineation resulting from one run of one method. Multiple samplings, if necessary, and level cuts of average links as well as multiple runs of LDA, AKM, and SLMA may help determine core clusters, and possibly continuous gateways between them. Limitations of this benchmark exercise should be kept in mind. It would benefit from tests on different reference datasets: any method can be trapped in particular data structures, and the results cannot be extrapolated without caution. As advocated below, processing multiple sources (lexical, citations, authors ...) and investigating the analogies and differences in their results will always prove rewarding. A number of in-depth benchmarking studies are found for hierarchical clustering (*Milligan* [2.129, 130] not covering the last techniques), discussing the generation of test data as well as comparisons of algorithms. For community detection, usually taken as a synonym of graph-based clustering rather than clustering of true social (actors) communities, [2.145] ranked first Infomap, then Louvain and Pott's model approach [2.161]. *Leskovec* et al. [2.146] studied the behavior of algorithms with increasing graph size. *Yang* and *Leskovek* [2.147] reflect on the principles of clustering outcomes compared to institutional classifications.

Bibliometric Mapping

Classical Way. Most classical bibliometric mapping, as well as information retrieval, relies on substantive (*feature*) representations of words, word combinations, citation, indexes, and so forth. Substantive representation implies legibility and interpretation by experts or users, and a condition for bibliometricians or sociologists to check and possibly deconstruct the document linkages. It contrasts with featureless machine representation applicable for example to distances of texts (see below). In contrast, the substantive approach is deepened in semantic studies: ontologies and semantic networks suppose more elaborate investigation of term relationships. Bibliometric mapping and IR techniques are both a client of ready-made semantic resources, and providers of studies, supported by data analyses, likely to help the construction of thesauri and ontologies.

The standard bibliometric model starts from the data structure of articles, essentially a series of basic article $\times$ attributes matrices, one of these reflexive: article $\times$ cited references, where references can also stand as attributes. The derived article $\times$ article matrices (e.g., bibliographic coupling, lexical coupling) and elements $\times$ elements matrices (e.g., coword or profiles, cocitation or profiles) cover a wide range of needs. Clusters of words are candidates for conceptual representation, concepts which in turn can index the documents. Likewise, clustering of cited articles reveal intellectual structures and in turn index the citing universe. Basically, the attributes (words from title, abstract, full text; keywords list, indexes—other fields like authoring) are processed in bags of monoterms or multiterms, recognized expressions or word *n*-grams. Standard bibliometric treatments rarely go further, semantic studies do, for example by using chain modeling of the texts. All these forms allow for control and interpretation of linguistic information.

Assuming that the final purpose is to classify or delineate literature, the access is dual: direct classification of articles after their profile on the structuring elements (words, cited references), or a detour by the structuring items: word profile (especially coword), citation profile (cocitation), index (or class profile) including coclassification, when applicable. The basics of citation-based mapping were established in the 1960s and the 1970s: bibliographic coupling [2.109], chained citations [2.162], cocitation [2.101, 163], author cocitation [2.164], coclassification, etc. The lexical counterpart, with its first technical foundations in *Salton's* pioneer works [2.165], was reinvested by English and French social constructivism in the 1980s [2.166–168] with a stress on local network measures quite in line with the development of social network analysis in that period [2.169]. In bibliometrics, the true metric approach of text-based classification, Benzécri's correspondence analysis [2.134], remained confidential. For convenience reasons, many large-scale classifications relied on proximity indexes and MDS or hierarchical single-linkage (ISI cocitation). We return later to word–citation comparison and combination (Sect. 2.4).

Developments. The principles above, mutatis mutandis, are kept in further developments of citation mapping: the approach through citation exchanges, mentioned in Sect. 2.2, assumes predefined entities, journals for example. At the article level, symmetrical linkages between articles, or between structuring elements, are classical: large-scale cocitation (*CiteSpace* [2.170]). *Glänzel* and *Czerwon* [2.171] advocated bibliographic coupling. As already mentioned, direct citation linkage clustering, the first benchmark for cocitation and coupling in *Small's* princeps paper [2.101], is considered as particularly able to reflect long-period phenomena [2.92, 172, 173] but not short-term evolutions. It turns out that the time range picked and the granularity of groupings desired might suggest the choice between the three families of citation methods to reflect structure and changes in science.

From the theoretical point of view, cocitation (respectively coword) is semantically superior to coupling, by visualizing the structure of the intellectual (cognitive) base, but requires a secondary assignment of current citing literature. Coupling as such, because it by default spares the dual analysis (the cited structure; the lexical content), is semantically poor but bibliographic coupling handles immediacy better than cocitation does. However, this depends on the computer constraints and the settings: the thresholding unavoidable in cocitation analysis drastically reduces weak signals that are accounted for in coupling. The dependence of the maximum retrieval on the threshold of citation and the assignment strength (number of references), in a close field, is modeled in [2.59]. Quite a few authors compared the methods empirically [2.128] over a short time range, [2.173, 174]. These studies are not always themselves comparable in their criteria, nor are they convergent in their outcome, so that it is difficult to come to a conclusion on this basis alone.

The new data analysis toolbox (fast graph unfolding, topic modeling) gradually pervades large-scale studies. From the domain delineation perspective, a general answer in terms of single best cannot be expected. The benchmark above reminds us that classical methods, apparently outdated in the big data era, still prove to perform quite well. Let us recall a few issues in clustering/mapping for bibliometric purposes, especially delineation.

A few Clustering/Mapping Issues

As other decision-support tools, maps in bibliometrics receive contrasted interpretations. In a social constructivist view, maps are mainly viewed as tools of stimulation of sociocognitive analysis and also as supports of negotiation with/amongst actors. If technicalities are not privileged, there is clear preference for local network maps, preferably lexical or actors-based, connected to sociocognitive thinking. Bibliometricians and librarians are keener on quantitative properties and retrieval performances. Expectations as to ergonomics, granularity, robustness, clusters properties, and semantic depth, largely vary depending on the type of study.

Ergonomics. Map usage benefits from new displays with interaction facilities. A tremendous variety of mapping methods is available ([2.175] although in practice a few efficient solutions prevail). The progress in interfaces (scale zooms, bridges between attributes, interaction with users...) changed the landscape of mapping. If adding cluster features to cluster maps is trivial [2.176], the systematization of overlay maps by *Leydesdorff* and *Rafols* [2.177] is quite appealing. Since delineation tasks often deal with multidisciplinarity, multiassignments, and cluster expansion, various types of cross-representations (Sect. 2.4) including overlay maps are quite convenient tools for discussion.

Granularity. The granularity considered here is the smallest unit handled. Progress of data analysis allows large-scale work with a fine granularity. Document-level maps are now regularly proposed by *Boyack* and *Klavans* [2.91]. The classical alternative in bibliometrics uses the *journal* molecule instead of publications, with the advantages and shortcomings already discussed. Delineation tasks used to be conducted at the journal level and this convenient solution can be somewhat improved using a core–periphery scheme with multidisciplinary qualification [2.178]. The interest of journal granularity for delineation remains dependent on the specialization profile at the scale considered, so is quite field-dependent. The best fit to the journal approach is found in fields with a strong editorial focus, such as Astrophysics, but [2.179] recalls that the general rule is the superiority of document granularity. At the global science level, journals or even journal categories are an option for sketching great regions [2.177], with low precision ambitions. In favor of journals, their persistence as institutional entities with slow demography, facilitates longitudinal approaches, again at the expense of precision (Sect. 2.3.2, *Dynamic Clustering*). Granularity does not reduce to the question of journals versus document level. It can also suggest methodological choices, e.g., the family of citation method to select, depending on the objective, taxonomies of disciplines or finer level research fronts in a broad sense.

Shape/Properties of Clusters. Ex post supervision of clusters (built by unsupervised methods) is a critical stage of studies. Discussion on the cluster aggregate features, or sampled articles, is much easier if clusters are reasonably homogenous. Therefore, the ability to recover clusters of any shape (elongated, nonconvex...), which is essential in other contexts (say image analysis), may not be desirable in bibliometric mapping. A few strongly linked compact clusters is easier to assess than the equivalent elongated class. The skewness of cluster distribution is another concern, especially in citation clustering, and the inflation of microclusters with poor connections is inconvenient—an argument voiced in favor of a direct citation approach for high-level taxonomies. From this point of view, the slight tendency of average linkage towards homogeneity and the tendency of k-means towards size balance, giving a moderately skewed distribution of cluster size, may be seen as *desirable biases* (refer to [2.146] in the context of community detection) with respect to further cluster supervision. As the benchmark exercise has shown, this does not prevent average linkage from recovering heterogeneous structures.

Soft Versus Hard Clusters. For reasons of convenience and computer efficiency, hard clustering is widespread but remains a violent approximation of the complexity and intrication of community networks and semantic relations in scientific literature. Hard clustering is sometimes the first stage of a two-stage classification: Cocitation analysis usually combines hard clustering for cores in the cited universe, and assignment of the citing literature is tantamount to soft clustering of research fronts. Reciprocally, starting from hard bibliographic coupling clusters makes it possible to generate a soft image of cited clusters. The conditions of assignment parameters in the second stage determine the degree of overlap. This is true also for factor analyses more suitable for overlapping entities, especially with oblique factors, i. e., principal axes of clusters upon which any entity, in or out, has a projection. The query expansion or bibliometric expansion practiced at the cluster level also builds soft clusters from an existing hard partition on the same data, therefore enhancing the recall at the cluster level. More generally, the wide development of probabilistic clustering is consistent with fuzzy approaches of assignment of particular articles/items.

Multilevel visualization of partitions is valuable for discussing topic or domain borders, especially when obtained from techniques which do not favor cluster homogeneity, or exploring strongly multidisciplinary phenomena. For example, assuming a strong proximity of two topics A and B, it is interesting to know whether this proximity is localized—say to subclusters A1 and B1—or distributed. Local intense linkages may prefigure capture of a subcomponent or merge A1–B1. Such interpretation only makes sense with robust methodology.

In a cluster selection process for delineation, all things being equal, soft or fuzzy clusters are allowed to extend towards shared areas, and then slanted towards recall at the cluster level. This applies to the boundary clusters, with an effect on a domain's delineation. However, bibliometric use of soft clustering remains limited and does not usually depart from the holistic perspective (Sect. 2.3.2, *Semantics, Statistics, Informatics*).

Robustness and Evaluation Issues. Robustness is an essential aspect of data analysis applied to bibliometrics. Sensitivity to data issues, to the type of network, to metrics and clustering algorithms, lead to rather different solutions. Ground truth or even gold standards are generally unavailable. In empirical studies, analysts have to get along both with biased representation of panels and divergences of techniques, as well as sensitivity to settings within one technique. We already mentioned general problems of bibliometric data, especially coverage. Within a given data corpus, the skewness of informetric distributions is a powerful foundation of robustness, but many sources of instability remain. The particular question of time robustness is sketched later.

Sensitivity to the Network Weighting and Metrics. For memory's sake, some prior transformation of bibliometric networks is practised to compensate across-domain differences, such as citing behavior. In such case, the value of linkages are weighted by a function of the number of inlinks of given groups (tantamount to classical cited-side normalization) or the number of outlinks. The latter is present both in influence measures (*Pinski* and *Narin* [2.44], revival in the last decade [2.180]) and the limit case of citing-side normalization which presents original properties [2.181, 182]. Citing-side normalization of the citation network is a limit case (removing iteration) of *Pinski* and *Narin* influence weights [2.44]. It is strictly classification-free if the basic normalization unit is the paper or the journal [2.181]. It exhibits interesting properties for any basic unit making sense, e.g., domains: the dispersion of domains' impacts calculated this way with normalization at the domain level is a measure of interdisciplinarity of science in a steady state system [2.183].

A major native characteristic of bibliometric networks is the skewness of node degree distribution and resulting polarization: citations, Zipf–Mandelbrot word usage, Bradford concentration—in connection with concentration generating models recalled above in social network theory. Concentration gives tremendous selective power and at the same time, calls for corrections in IR context for information retrieval and usage, depending on the context. A vast choice of metrics or quasimetrics (similarity indexes) is available, introducing weightings with some inverse function of frequency, especially useful in a mapping context. It is common knowledge that various similarity indexes produce contrasted perspectives. Coword analysis pioneers, notably, compared the unweighted index (raw), the asymmetrical (inclusion) index, the partially weighted index (Jaccard, Ochiai among others), the strongly weighted index (p-index or affinity amenable to a similarity). After thresholding, the landscape of the transformed networks is quite different: the first two indexes tend to keep the frequent items as hubs, the last one highlights infrequent words and associations at some risk of overexposure of rare forms, amongst them typing errors.

Analogous normalizations, from the abundant repertoire of similarity indexes, are frequent for cocitation [2.184] and coauthorship analysis [2.185, 186].

Clustering algorithms build on the final network in various ways. Obviously, any delineation based on such weighted networks of structuring elements—where skew distribution is the rule—will be quite sensitive to methodology. In bibliometrics, the contrast is extreme between steep landscapes generated by raw measures, dominated by the centrality of hubs, and information-driven strongly corrected configurations, at the risk of instability and errors on very low frequencies. Intermediary options are often picked, for example Ochiai–Salton and Jaccard measure. Document coupling relations, similarly, depend on the normalization of term frequency, typically inverse frequency weighting, Hellinger, etc. built-in or not in data analysis methods (TF-IDF, χ^2 in correspondence analysis, etc.).

Asymmetrical Relations. Specific to citations, a complete model of citation exchanges requires some native or constructed aggregation with relatively stable entities (authors, journals, pre-existing categories, etc.) in order to allow both in- and out-linkages while document-level direct citation is unidirectional—with exceptions. Asymmetry at the journal level inspired the CHI classification of journals after their theoretical versus applied orientation [2.44] on the hypothesis that applied science journals tend to import knowledge and export citations, and reciprocally for basic science journals. The same phenomenon appears at the field level (cell biology versus medical research, for example).

The valuation of bilateral relations calls for methodological choices which can largely affect mapping and delineation. Take the simplest case where i and j denote two aggregates (journals, domains...) and assume the ij link is normalized on the basis of the total outflow of i and the total inflow of j, and conversely for the ji link. Let us calculate the bilateral link between i and j by the arithmetic mean, the geometric mean, and the maximum of these two unidirectional normalized flows, a simplified variant of [2.87, 187] for the sake of the example. Should these valued networks be used for delineation purposes, they would tend to produce rather different results. The multiplicative indexes trivially penalize one-way relations typical of vertical channels, and tend to group entities with balanced relations, either particularly integrated channels or basic science fields with multidisciplinarity relations, or else clients sharing methods or products. In contrast, the maximum index tends to retrieve vertical channels (say cell biology–medical research) regardless of flows dissymmetry. Additive indexes stand in intermediary position, and appear as a middle-ground choice.

Semantics, Statistics, Informatics. Scientific domains at the mesolevel represent a considerable amount of data, especially in longitudinal series. The computing requirements, even with sparse bibliometric matrixes, are high, driving towards clustering or spectral analysis algorithms with high efficiency. The trade-off between computer efficiency and semantic power is far from simple. Correspondence analysis [2.134] was amongst the first factor technique to exhibit some semantic power in textual applications, especially a robust capability to group quasisynonyms with the distributional equivalence property. In its wake, postfactor analyses keep claiming some semantic power (Sect. 2.3.2) and built-in mapping capability. In parallel, local similarity techniques associated with traditional or innovative clustering methods from network analysis privilege the native graph of proximity and elements/links groupings. In those approaches the duality (structuring elements $\times$ documents) needs assignment decisions (e.g., research front assigned to cocited core) with a semantic dissymmetry as to the internal scrutiny of clusters: while the detailed map of structuring elements is appealing for cluster evaluation (cited cores; within cluster word-map), the document coupling map, internal to a cluster, is hardly interpretable alone as stressed before.

Now, if word-maps present high potential for sociological interpretation, mere lexical associations remain semantically shallow with regard to truly semantic analyses. A common limitation to all these methods is the *bag of words* overlooking the rank of words and the structure of statements—the downside partly alleviated by multiterm treatment (noun phrases). Citations present a fuzzier relation to semantics (Sect. 2.4) but cocitation cores are nevertheless understandable for experts. Labels or lists of descriptors directly issued from cocitation or coword cores, for example a ranked list of specific terms, or indirectly rebuilt from clusters obtained by coupling, are common but limited auxiliaries for evaluating clusters. Cards might be reshuffled with new competitors to LSA such as neural word embeddings (Sect. 2.3.2). In addition to the similarity calculations in the word–context, useful for information retrieval, semantic calculations on word vectors are possible, allowing good performance in analogy tests (i. e., "Find X so as X is to A what B is to C") or inference operations on these vectors, such as `king` $-$ `man` $+$ `woman` $\rightarrow$ `queen`. This gain in semantic precision suggests that, applied to scientific corpora—now increasingly available in full text—it could allow in the future for an analyst to select the semantic dimensions relevant for delineating scientific fields and constitute crisp or overlapping groups of articles (or parts of these) in this subspace.

A recurrent problem of more traditional bibliometric representations, a counterpart of statistical simplicity and computer efficiency, is the holistic character of linkages, especially if combined with hard clustering. In document coupling techniques, either word-based or citation-based, the standard linkage measure is the weighted and normalized number of words shared. In lexical coupling, an implicit hypothesis is that the (weighted-normalized) number of shared tokens reflects the dominant semantic dimensions of the paper. For example, if very few words or references refer to methodology, this dimension will contribute less, all things being equal, to the shaping of bibliometric similarity, which can be misleading. In the opposite case, if methodology markers prevail, a transdisciplinary corpus will tend to be split between hard science literature and soft science literature on the domain, whereas mixed clusters would probably reflect the domain structure in a better way. Should the linkage between two clusters need explanation, this should be inferred from the features and given the titles of the two clusters, unless the technique includes indicators of contribution. In clusters of structuring elements (word graphs, cocitation cores) the relations are interpretable when zooming in on the fine-grained networks of words or cited articles, but without semantic characterization.

In a delineation context, a minimum of semantic break-up would make the scrutiny of the border region easier and faster. It could especially orient discussions on preferential extensions of a core zone towards neighbor clusters with shared methodology but new objects, shared object with new methods, etc. Ad hoc simple characterization of vocabulary has been successfully applied for other purposes, e.g., the level of application of biomedical research journals [2.188]. However, manual semantic tagging is quite intensive and field-specific. At the document level, many natural sciences articles can be labeled with simple semantic combinations. In computational linguistics, many works since *Teufel* et al. [2.189] (argumentative zoning) address this issue of categorization of scientific discourses and automatic annotation, applicable for example to the summarization of scientific texts. Several proposals on categorization of arguments have been made, many of them at the experimental stage. *Liakata* et al. [2.190] developed and automatized the core scientific concept (CoreSC) categorization whose first layer distinguishes 11 categories: objective (hypothesis, goal, motivation, object), approach (method, model, experiment), and outcome (observation, result, conclusion). This line of research is extremely promising for bibliometric studies, especially domain delineation, but remains for the time being limited to small universes. In the meantime, oversimplified semantic indexing would help a lot in qualifying interdocument or intercluster relations. Figure 2.4 shows a fictitious configuration where documents are naively described by semantic triplets with various degrees of kinship. The graph display could be replaced by a superimposition of three partitions, each one upon a different semantic dimension.

More intensive semantic mapping relies on sophisticated ontologies, knowledge models, and semantic networks. If such resources have not been established beforehand and published, bibliometric studies cannot generally afford such heavy developments, however see [2.191].

Directly opposed to semantic approaches are non-feature methods from computer science, which ignore the substantive representations and even more so the semantic content. In various IR/bibliometric applications (disambiguation of authors and affiliations, proximity of documents, detection of plagiarism) similarity between texts may be calculated on the basis of character n-grams [2.192] rather than *feature* word n-grams which is somewhat standard. The link to the minimal unit with semantic load, the word, is lost (almost completely for low values of n). The usual metrics can be applied to n-grams. A more radical way using the bit sequence representation with further compression, is the basis of measures like normal compression distance (NCD in [2.193]). NCD is a dissimilarity measure which is an approximation of the general Kolmogorov information distance [2.194, 195], parametrized by the compression algorithm. A *normal* compressor should satisfy four properties:

1. Idempotence
2. Monotonicity
3. Symmetry
4. Distributivity.

From the linguistic point of view the compression method is a black box. It nevertheless exhibits rather good performances for calculating text similarity with a most indirect semantic power of forms unification. The normalized Google distance (NGD in [2.196]) is the transposition to Google searches, at the word level, of the NCD, keeping the *feature* characteristics of the coword analysis and its semantic power. Its native application builds on lexical associations from millions of users.

Table 2.3 summarizes the degree of semantic ambition in the case of lexical approaches—transposable to citation attributes.

Dynamic Clustering. The delineation process has to face changes in the configuration of networks [2.124], affecting the value of a delineation solution at a particular moment. Dynamic clustering is understood in two (related) acceptations.

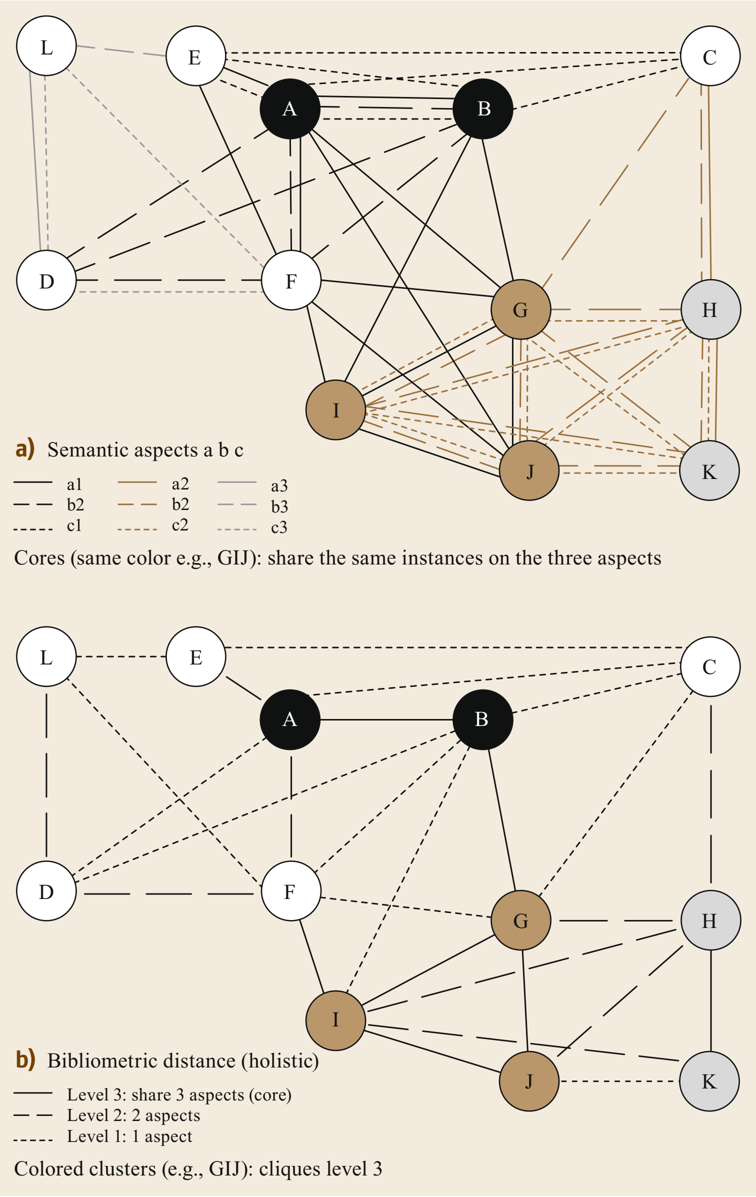


Fig. 2.4 Semantic and bibliometric linkages. This figure sketches bibliometric holistic distance versus decomposition into semantic links, with the (heroic) hypothesis of tagging with only three criteria, e.g., a = theory–hypothesis, b = experimental method, c = observation–test. For example a1, a2, a3 on the figure denote different hypotheses. The second panel represents three kinds of semantic relations. An article is described by a triplet a, b, c. For example, the documents G, I, and J are described by the same triplet {a1, b2, c2}. Documents G and I, for example, are connected by three links. The second panel aggregates information in a single type of linkages with varying degree of intensity. Here the bibliometric linkage is assumed proportional to the number of shared semantic instances, which is of course arbitrary. In the real bibliometric world, the lexical coupling linkage heavily depends on the most developed aspect(s)

Table 2.3 Semantic interpretation potential of various approaches

Structural items metrics	Document metrics (required in delineation task)	Category	Semantic interpretation versus black-box
Semantic network	Indirect through indexing/ assignment to word structures	Feature	Strong, feature
Word profile/coword	Indirect through indexing/ assignment to word clusters	Feature	Light, direct
Document profile/coupling	Direct: lexical coupling	Feature	Indirect, through indexing/assignment/labeling
–	Direct: char *n*-gram proximity	Featureless	Black-box[a]
–	Compression distance	Featureless and global	Black-box[a]

[a] Clusters of documents based on nonfeature proximity can be interpreted by going back to substantive elements, e.g., their word profile.

A first point of view is the adaptation of algorithms—and computer resources—to processing massive data streams, typically texts, an example today being online social networks. The initial k-means algorithm of *MacQueen* [2.197] was already an *online* incremental one, generating a cluster structure in one pass over the dataset—the usual iterative version, which converges to a solution independent of the presentation order of the data vectors, is due to *Forgy* [2.198]. Dynamic text stream mining is a growing topic in the machine learning and big data mining research communities. Changes in the cluster structure may reflect algorithmic artifacts as well as real phenomenona, hence ideal methodological characteristics are a) global optimum seeking and b) insensitivity to data ordering. An example of an incremental hierarchical clustering method for texts is [2.199], and a frequent itemsets dynamical clustering example is [2.200].

A second point of view focuses on the domains/topics picture and their description over time, through cluster time series, including the issue of time robustness in one-shot pictures. Again the distinction between clustering/mapping on *structuring elements* (e.g., cocited articles or lexical relations) and *direct clustering of literature* (e.g., bibliographic or lexical coupling), in techniques privileging classification in one space matters. The first family offers solutions with some durability. The repertoire of words gradually evolves. Change of the intellectual repertoire of cited literature, subject to an aging process, is usually faster but, except in emergent or revolutionary fields and in intrinsically rapid ones (e.g., computer science), it respects a mix of new and old literature. This gives some clue of robustness, in the short term, to the cluster solutions. By and large, in slow evolution processes, information cores are more persistent than peripheries. In one-shot clustering, working on pluri-annual window data reinforce the robustness of the breakdown and permit the cross-characterization of novelty (median of the cocited core) and internal growth in the span of the window (average date of front) [2.176]. Characterizing fine granularity hotspots in the network, such as local preferential attachment processes, may help to spot promising weak signals. Taxonomic applications of direct citation linkages might still benefit more from long time window settings. This would sketch, as noted earlier, a possible trend towards division of tasks between direct citation, cocitation, and bibliographic coupling in function of targeted granularity and immediacy of results.

By construction, direct clustering of documents over a time period (say the year) favors immediacy, but is not prolongable without a detour by the structuring elements and derived cluster labels. Another way consists in picking a coarser granularity, especially the journal level, at the expense of a heavy loss of precision. Short-time changes may be addressed by projecting a solution for a period on the reference solution of another period, a classical process in factor analysis applicable to other methods; an early example within bibliometrics is found in *Noyons* and *van Raan* [2.32].

A delineation process of any kind may be run on successive slices of time [2.201] of different lengths, with or without rolling averaging filters. A dynamic variant of LDA is [2.202], in which the word distributions of each topic varies in each time slice, where the number of clusters is fixed. Interesting historiographic insights accounting for cluster demography (emergence, death, splitting, merging...) are exhibited by longitudinal chaining of clusters, known since ISI's Atlas of Science, see [2.203–206]. The latter work is based on lexical series. The predictive value of such series, along with life-cycle models, remains a quite difficult issue.

Last but not least, the rendering of change is closely linked to dynamic models of science where structure emerges from local properties, for example in the preferential attachment model. In this view, over time, breakthroughs (scientific or technological) shape the citation profiles of followers, a common mechanism in (co)citation bibliometrics. Local accretions around hot papers are amongst the signs of emergence. The symmetrical question over whether the referencing (or lexical) profile of papers has some predictive value, remains open. This connects to the controversies about interdisciplinary distal transfers in the discovery process, quoted above, which echo the combinatory nature of invention and innovation stressed by Schumpeter. The intuitive but bold hypothesis stating that the more distant the knowledge transfer, the more radical the discovery or invention is, nevertheless, tricky to test (definition of scientific or technological distance from models A or B–C, scale issues). Attempts to characterize scientific breakthrough and radical inventions, with an ex ante notion, are found for example in [2.207], using both citations and patent classification [2.208], using changes in forwards and backwards citation profiles [2.209], using citation contexts of outstanding discoveries.

2.3.3 Conclusion

By and large, bibliometric mapping provides landscapes with aggregate groups (clusters; local factors, etc.) likely to be assessed, and implementation of multistep and cross points of views help to distinguish cores and border regions, the latter calling for cluster evaluation, see Sect. 2.5.2. No mapping method is superior on all criteria and many factors are at play: the bulk of data, the type of network, the nature of the problem, and the ergonomics of outcomes for an easy supervision. IR search remains an alternative or a valuable complement to mapping. The next section focuses on hybrid techniques.

2.4 Multiple Networks and Hybridization

This section addresses the multinetwork approaches. We shall especially develop the combination of textual and citation networks but most types of bibliometric (and altmetric) networks can naturally contribute where appropriate. The forms of hybridization encompass a wide scope from fully integrated approaches to parallel schemes aiming at comparison and eventual combination, with intermediate sequential schemes.

2.4.1 Multiple Networks

A given document may be accessed by search strategies pointing at all searchable fields of data or metadata. Modern IR, going beyond the direct query–document similarity, integrates, with the cluster hypothesis and later the cognitive model, the documents' multiple spaces and networks, including citations and collaborations. Bridges between lexical and citation universes were built, especially for labeling purposes (e.g., keyword-plus [2.210]).

Likewise, major streams of study in the sociology of science have coined general theories accounting for the various manifestations of scientists behavior in communities: communication, collaboration, publication, rhetoric, citation, evaluation. The networks of science, although diverse, originate in the same ground. As a result, many classes of bibliometric questions (topic identification, characterization of emergence, static and dynamic mapping, diffusion processes, knowledge flows in science and more generally in the science–technology–innovation system) can be answered by working on different networks, with respect to their specificity. The multinetwork approach to bibliometrics, both in terms of comparison and complementarity, appears as a natural mode of thought.

With the coming of age of data representation models such as entity–relationship for relational database management system (RDBMS) implementation and of network analysis methods, IR scholars and bibliometricians in the early 1990s found flexible tools for easy handling of different dimensions of publication data. In the last decades, the culture of data mining encouraged mixes between several networks for pragmatic purposes [2.211]. We recall the key role of author networks (Sect. 2.4.2) before focusing on text and citation networks (Sect. 2.4.3) and finally their hybridization (Sect. 2.4.4).

2.4.2 Networks of Actors

The first analyses of scientific communities in the 1970s led to some disappointing results as to the unambiguous assignment of particular scientists to a particular group. In a short history of domain delineation *Gläser* et al. [2.26] recall among others *Mulkay* et al.'s work [2.9] and Verspagen and Werker findings [2.212]. The archetype is the coauthorship graph. *Price* and *Beaver* [2.18], *Beaver* and *Rosen* [2.213], *Luukkonen* et al. [2.214], *Kretschmer* [2.215], and *Katz* and *Martin* [2.216] laid the first layers of collaboration studies in connection with invisible colleges. Author-based models of science are amongst the central topics in science studies and bibliometrics. Studies on scientific collaboration are out of the scope of this work, but let us recall the macrolevel studies of the determinants of cooperation in the wake of *Luukkonen* et al. [2.185], geographic proximity [2.217, 218], cultural links [2.186], and individual/collective behavior [2.219]. Those studies emphasize the importance of metrics and normalization in the interpretation. At the microlevel, proposals for mechanisms explaining the structure and dynamics of social networks were recalled in Sect. 2.3.

Networks of actors present a major theoretical interest: they stand at the crossroads of actual social networks' mathematical modeling and sociology of research, and bridge invisible colleges with cognitive structures [2.220]. They also show some drawbacks, echoing the scholars' disappointment noted above. Communities detection in practice faces the issue of names unification. For a long time, the problem has been both terribly costly and time consuming for data producers and bibliometricians, at both the institutional level and the author level, as stressed again in the name game project APE-INV (Academic Patenting in Europe Project), e.g., [2.221]. Great progress is ongoing due to the ORCID (Open Researcher and Contributor IDentifier; with the unique identifier of researchers), ISNI (International Standard Name Identifier), and GRID (Global Research Identifier Database) initiatives among others.

Another issue, especially for small topics detection, is the width of the competence spectrum of productive authors likely to produce some noise, but this shortcoming is alleviated at the level of large domains. In this case perhaps, community detection (in a narrow sense) has arguments to compete with citation or lexical clustering. However, in most practical studies multiscale vision is required: not only does the target domain matter, but also the subdomains. At this scale, the polyvalence of authors limits precision. The problem may be reduced by time-restriction filters, the link-level technique, external information, or hybridization with citation or word information. Similar issues appear in *author cocitation* versus *article cocitation* [2.164, 222]. Author cocitation opened insights in the study of invisible colleges, with connection to researchers' sociology. Topics mapping as such is better addressed by document-level cocitation.

The interplay of coauthorship, citation, and linguistic networks as a mirror of sociocognitive activity is increasingly gaining attention: relations between contents and actors' positions [2.223, 224], between citations and coauthorship, and any or both of these with texts [2.220]. Is the multiple approach a step towards more powerful models of authors and community behavior, able to unify the diverse representations? This unification would spread benefits over bibliometric analysis, including delineation tasks. Nonfeature methods have not waited for unification (see below) to mix up all types of information, but they sacrifice the substantive depth of analysis.

However, the quest for unification might be hindered by the specific features of every bibliometric network. Changing the type and parameters of the network is like observing the universe in various wavelengths. The most dense objects produce various forms of energy and tend to be retrieved albeit with diverse volume and appearance. Less-dense objects like clouds of various composition can be seen only in specific parts of the spectrum. Likewise, we may conjecture that dense and isolated objects will be retrieved from any network fit for precise analysis [2.113], especially words and citations and perhaps coauthorship clusters. Sociological investigation is expected to confirm such configurations as bounded invisible colleges. In less dense and more connected areas, each network is likely to produce nonsuperimposable images, with different sensibilities. The convergences suggest strong forms with easy sociocognitive interpretation, while the divergences ask for careful tests and investigation. The sociology of translation associated less dense areas to emergence or ultimate evaporation phases.

2.4.3 Citations and Words

Lexical and citation characterization classically used in bibliometrics are appropriate for clustering of themes and mapping at various scales, on the basis of the toolbox sketched in Sect. 2.3.

A few Analogies and Differences

General. One difference naturally lies in the nature of the original relation: direct attributes for linguistic elements, reflexive interarticles for citations, with several consequences. Firstly, the granularity: words are an ultimate attribute (in classical *feature* methods) whereas cites target the full article semantic aggregate. Then, the linguistic content of citations is not explicit, and requires a statistical detour via the text fields and the data model, to emerge (automatic labeling of clusters with their specific vocabulary, citation contexts). Secondly, the time relation, not explicit in lexical relations, directly appears in the citation link, both cited and citing article being dated. Bibliometrics makes a large use of this diachronic relation in immediacy–aging studies. In contrast, the word content of an article is readily legible, but deprived of temporal information beyond the article date of submission/publication. Going further requires statistical studies to date the word in terms of chronological profile of use. Longitudinal studies on words have to rely on time statistics of use, typically with the assumption of achronicity: constant meaning over time. This is a bold statement in some cases. Beyond classical dating of word or word linkages after their usage, determined by the obsolescence of topics, natural language analysis paved the way for analyses of word transformations in a scientific context [2.225].

With respect to these constraints, a large class of bibliometric, IR, or altmetrics issues can be addressed by the lexical method or the (generalized) citation method with the exception of specific direct chaining [2.162]. Symmetrized relations (cocitation, coupling) mitigate the diachronicity, albeit underlying time features can be invoked if required. The reformulation of the dynamic chaining research fronts [2.205] is emulated by word-based clusters [2.202, 206]. Only the former directly contains citing–cited information for immediacy characterization.

Because of limitations (indexer effect) and lack of reactivity of controlled language, modern bibliometrics moved gradually towards natural language, building on the increasing availability of full text resources and lexical treatment. In spite of progress in computational linguistics, the NLP remains tricky, a counterpart of language richness and versatility. Polysemy, metonymy, synonymy, figures of speech, metaphors, acronyms, and disciplinary jargon are well-known linguistic traps of linguistic difficulties that users, bibliometricians, and retrieval specialists have to cope with. Unification (stemming and lemmatization, synonymy detection) also benefits from clustering techniques. Unsupervised homonymy tracking is a more challenging problem, since bridges in word clusters may be rooted in concept transfers or polysemy or else simple homonymy. This issue is somewhat alleviated in small (narrow context) studies. If elaborate ontology or semantic networks are seldom off-the-shelf, useful tools for term extraction, parsing, and coword exploration are available. Stemmers (with *Porter*'s stemmer milestone [2.226]) or, a step further, lemmatizers are efficient with some risk in precision. New massive techniques, such as the above-mentioned deep learning-based or neural networks or targeted methods such as neural word embeddings, might bypass or alleviate costly preprocessing. Constraints of bibliometric studies dealing with large data universes are usually incompatible with refined semantic treatments, but the supply of large-scale statistical semantics resources might spare costly ad hoc developments. We mentioned (Sect. 2.2.4) a possible revival of controlled vocabulary supported by bibliometric treatments.

Statistical Background. The common feature is the skewness of frequency distribution found, among other disciplines, in information processes (Bradford–Lotka–Zipf trilogy, see [2.227]). The classical model to fit word distributions is the hyperbolic Zipf–Mandelbrot model. Other Paretian distributions are also used for citation frequency analogous to node degrees in the native oriented graph of citations. Similar skewed distributions are found in authors' collaboration graphs, with a distinction between scale-free distributions and small-world distributions (Sect. 2.3). The parameters of citation distributions are modulated by the citation windows, the parameters of word distribution are modulated by the type of lexical sources (title, abstract, full text...), the type of lexical unit picked, the language, and the richness of vocabulary.

Comparing the distributions of citations and words on the same corpus, some authors found that the latter appears more concentrated and less *complex* [2.33], thus less favorable in principle to precision—without forgetting the different granularity. Frequency weighting of linkages of the native word or citation networks, or similarity indexes with various types and degrees of normalization, may be implemented for retrieval or mapping purposes, for favoring information-rich elements in low and/or medium frequency. The precision of citation approaches was underlined in comparative retrieval tests, and especially the interest of cross-retrieval [2.228, 229]. As to co-occurrences, coword matrices tend to be less sparse but noisier than cocitations relations.

For the delineation work, the distribution of words or citations designs the background, with implications for interpretation, but what directly matters is the arrangement of documents after their texts or their bibliography. For this purpose, the typical approaches are the direct profile proximity on either type of structuring elements, words or references (*coupling* rationale or profile metrics in vector space), or the secondary assignment on prior classes of structuring elements such as coword, cocitation, or corresponding profiles. The distribution of node degrees in bibliographic coupling tends to be less skewed than in the original citation graph. Again normalization of distances or similarity by some function of inverse frequency can reduce the unevenness. The recall advantage of word-based techniques suggested their use in the large-scale mapping of clusters defined, beforehand, by citations [2.91]. There is some evidence in the same direction for patent–publication relations. Composite word–citation metrics are addressed in Sect. 2.4.4. Technicalities involved in term unification are also different. As information tokens, references are less difficult to match than natural language elements. Keys on cited references reveal effectivity and improve with standardization of entries, with residual difficulties in particular cases like citation analysis of patents towards science.

Sociological Background. The textual contents of an article and its bibliography are both the results of authors' choice in their community context. Both involve an intricate mix of scientific and social aspects: words and cited references are community markers and reflect the sociability of invisible colleges. A large body of literature (refer to the review [2.230]) has been devoted to citation behavior, including *Cronin's* classic work [2.231]. Whatever their determinants can be, Merton's rewards, *Small's* symbolic beacons or concept symbols [2.232], *Gilbert's* persuasion tools [2.223] or Latourian interests, the references mainly point towards the thematic groups where founding fathers, gatekeepers, and potential partners are found, which matters in science mapping. On the textual side, rhetoric and jargon expressing community habits, in addition to general words voicing interests, rejoin focused scientific terms—especially specific multiterms with medium frequency—to define topics. A substantial amount of convergence between texts and citations is therefore expected when the delineation of topics and communities are at stake. Some degree of parallelism may be found between relatively high frequency expressions (after filtering of stop-words) and highly cited articles in generic knowledge and multidisciplinary linkages. The measured convergence depends on the information unit and is likely to increase with small lexical units of citation contexts (see below).

However, the question arose as to which network is the more appropriate for describing science, at a time (the 1980s) where citation evaluation, indexing, and mapping were gaining interest. The social constructivist stream and the actor network theory mentioned above (Sect. 2.2) favored the coword networks [2.166] against citations to represent knowledge on a background of actor's interests. Texts appeared abler to depict more completely *science in action* [2.233] especially in controversial areas where social and cognitive aspects are inseparable, while citations were supposed confined to the capture of *cold science* with delays and incompleteness. The delay argument alone is less convincing for bibliographic coupling. Typical cocitation *research fronts* rely on a high-pass filter on citation or cocitation scores, favoring old articles, to reduce the data volume. Bibliographic coupling often works on the whole reference lists, letting recent and less cited references play. A residual effect of the publication cycle of the citing side nevertheless subsists. Similar delays may also occur in the use of new words or expressions qualifying a scientific technique.

In its very realm, academic science, citation analysis encountered lasting problems in quite a few disciplines, especially in a fraction of SSH, because of citation sparsity, incomplete processes of internationalization, and lack of coverage in databases. This argument is somewhat weakened nowadays because of data source progress and changing behavior of scholars confronted with science globalization and bibliometric evaluation. Citation analysis proved an appealing tool, including for the borderlines of standard literature, for example transfer documents (guidelines and even magazines and newspapers) explored in biomedicine by *translational research* for improving health system services [2.234]. See also the EUSTM website at https://eutranslationalmedicine.org. As to the coverage of technology, the transposition of citation analysis to patents was revealed to be rather successful [2.235] competing with lexical approaches [2.236]. It nevertheless requires acquaintance with specific citation rules and behavior in patent systems. The Internet produces linkages with an exploitable analogy with citations, as the Google search engine has demonstrated in the wake of Pinski and Narin's influence weights.

Citations are not without their shortcomings, stressed in voluminous literature from various horizons; see *Bornmann* and *Daniel*'s aforementioned extensive review [2.230], and for the defense, mostly, see [2.51]. For the reason stated above, citation biases are somewhat less severe in mapping applications than in citation evaluation (impact, composite indexes) which concentrate controversies. Latourian citations or rare negative citations do not add much noise to cocitation topics. Other downsides are more serious. The bandwagon effect in citation behavior tends to create spurious cliques in native cocitation networks, possibly hindering the discriminating power of citation relations. The inflation of the number of references in authors' practice, which is a long-term trend [2.237], also brings noise to conventional citation clustering. The disciplinary insertion affects the number of references (*propensity to cite*) justifying citing-side normalization approaches mentioned above.

Albeit language-dependent, textual analysis is media-free, which is valuable in fields where academic sources with standard citation behavior are not sufficient. Topics peripheral to the academic mainstream, or demanding a mix of heterogeneous data may be confined to text-based delineation.

In cases where no differential data coverage issue is faced, differences may arise between these expressions of scientists' behavior, resulting in alternative breakdowns into topics, independently from statistical properties. The expectation is that citations, albeit in a blurred and biased way, are more capable of tracking the intellectual inheritance. A single difference in the semantic mix, for example different methodology on the same category of problem, will probably better discriminate amongst microcommunities than lexical analysis, at least as long as those microcommunities do not secrete specific terminology.

Let us turn towards limit cases, special forms of particularism, especially perhaps in SSH where intellectual traditions resist globalization. Words as well as citations would distinguish between schools of thought with opposing theories, strong community preference, and distinct jargon: say in postwar period marginalist versus Marxist economists. In contrast, if the linguistic repertoire is shared by the two communities while they diverge in the intellectual base, the outcomes of the two approaches will be different. The reverse can be true, with a common recognition of the intellectual base but divergent traditions in terminology, perhaps again for reasons of national tradition. Such configurations, relatively rare, limit the generality of the conjecture stated above about local convergence of bibliometric networks in zones with high-gradient borders. Most of the time, a set of clustered papers belonging to a strong overlap of a word-based cluster and a citation-based one may be considered as a strong form, in a rationale already present in the first comparisons by *McCain* [2.228] on term versus citation indexing. The cognitive overlaps between information types was a keypoint in Ingwersen's model mentioned above.

Empirical Comparisons

The cross-check of cluster contents is a run-of-the-mill operation. For example, the enhancement of cocitation coverage by two-step expansion could be controlled by lexical means [2.127]. A few specific comparisons of the two mapping approaches on the same data are found in the literature. The scale is therefore different (subareas rather than a large domain) but the method can be applied to an overset expected to contain the targeted domain, as seen before. In an extensive study of a few promising fields in the 2000s, using bibliometric mapping, *Noyons* et al. [2.238, 239] warned about the difference of concepts: publications and keywords and concluded they were "totally different structures".

Fig. 2.5a,b Archipelago display: Nanosciences **(a)** and Genomics **(b)**. Data: Reordered cross-tabulate matrix of axial K-means clusters respectively from bibliographic and lexical coupling 50×50). Relative overlap (z-axis) measured by the Ochiai index. Reordering: ranks on 1-dim MDS, making the diagonal accumulation showing the visual convergence between the two breakdowns apparent. The line is sinuous because of discrepancies between c-cluster versus w-cluster size distribution. The visual rendering suggests superclusters at a larger scale. In the nano figure, the area of nanotubes as a whole is retrieved by both methods, but with two different breakdowns and more discriminative power on the citation side (after [2.113]) ►

An opposite conclusion was reached by *Zitt* et al. [2.240] on nanosciences and *Laurens* et al. [2.241] on genomics, previously delineated as a whole by a hybrid sequence method. They implemented a more direct comparison scheme on clusters respectively from bibliographic coupling and lexical coupling (natural language, titles–abstracts), using the same axial k-means method (AKM). Cross-tabulate cluster overlaps [2.242, 243] were reordered, giving a quasilandscape with a heavy and narrow diagonal load (Fig. 2.5). This gives evidence of a fairly good convergence of lexical and citation solutions, also confirmed by direct indicators.

On their high-level maps, *Klavans* and *Boyack* [2.244] and *Leydesdorff* and *Rafols* [2.245] also observe a reasonable degree of convergence. More general comparisons of mapping methods including textual are found in [2.173, 246, 247]. A recent exercise of mapping comparing cluster methods is reported by *Velden* et al. [2.248]. Most experiments, however, lack a ground truth reference, and techniques presented as gold standards are disputable.

More generally, suppose we built clusters of documents from several origins: lexical coupling, bibliographic coupling, fronts from cocitation, author coupling, etc. Those various cluster solutions may be individually mapped. They can also be simultaneously represented using normalized overlaps between w-clusters, c-clusters, a-clusters, with appropriate metrics. Profiles distance may be required to overcome the zero overlap between hard clusters of the same family, say w-clusters. Resulting matrices are still quite small and amenable to MDS display.

The fact that the agreement between citation and lexical approaches is good but not complete brings one more argument in favor of complementarity. One thing to keep in mind: because of the imperfect optimization of reordering and choice of the article rather than sentences or narrow contexts as the lexical unit, the global convergence tends to be underestimated.

Complementarity

Complementarity, rather than competition, already inspired the *citations in context* researches, initiated in cocitation studies [2.249, 250] which are a natural space to connect referencing, intellectual base, and linguistic aspects. In a step further than linguistic labeling entities in (co)citation analysis, the studies of *citation in context* range from simple context visualization in citation engines to investigations in the dynamics of science. They tend to reinvest research in action, associating language and communities' life. The linguistic and semantic analysis of citation contexts contribute to topics such as the citation types or motives [2.251], the classification and cross-analysis of the contents of the citing or the cited documents [2.252], the fine-grained relation of citation contexts and abstract terms [2.253], the exploration of new dimensions of scientific texts [2.254]. Some of these advances influence citation techniques in return. An example is the improvement of cocitation accuracy [2.255, 256].

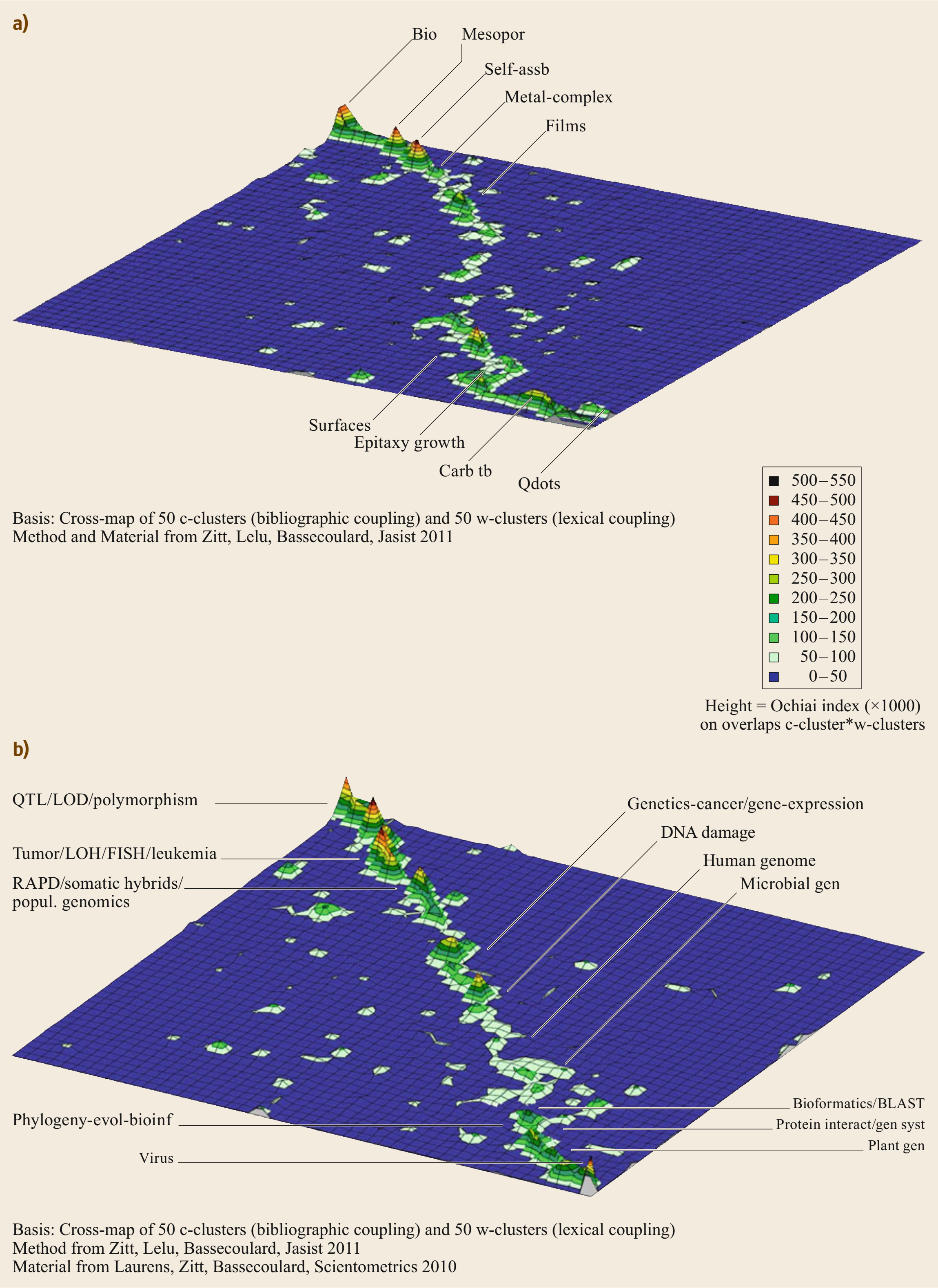
As a result of multinetwork or polyrepresentation hypotheses, some issues typical of one representation can receive a solution from the other. Convergence at the local level also creates spaces for complementarity: synonyms of any kind, for example, tend to be retrieved in the same citation-based clusters. Citation techniques escape linguistic polysemy and the reverse is true, but *citation homonymy* often due to matching keys, is a less important risk.

Finally, textual information preserves its advantages of availability, intuitiveness, and interpretation, with easy transposition to concepts and topics. A major shortcoming is the complexity and ambiguity of natural language, resulting in poor precision in the case of unsupervised protocols. In spite of the composite unit handled (the full article rather than the narrow concept), citations are appealing for tracking intellectual influences and often less noisy, at the expense of lower recall in weak signal configurations.

The capability of pure lexical approaches to emulate citation-based or hybrid approaches in challenging topics such as the aforementioned description/anticipation of early stages of domain emergence, remains a challenge.

2.4.4 Hybridization Modes

Looking for optimal exploitation of these contrasting properties is the quest of hybrid techniques, in line with pragmatic mixes of dimensions in IR-type delineation for bibliometric purposes. The same pragmatism inspired mixed information classification of web


a)
Bio
Mesopor
Self-assb
Metal-complex
Films
Surfaces
Epitaxy growth
Carb tb
Qdots
500–550
450–500
400–450
350–400
300–350
250–300
200–250
150–200
100–150
50–100
0–50
Height = Ochiai index (×1000)
on overlaps c-cluster*w-clusters
Basis: Cross-map of 50 c-clusters (bibliographic coupling) and 50 w-clusters (lexical coupling)
Method and Material from Zitt, Lelu, Bassecoulard, Jasist 2011
b)
QTL/LOD/polymorphism
Tumor/LOH/FISH/leukemia
RAPD/somatic hybrids/
popul. genomics
Genetics-cancer/gene-expression
DNA damage
Human genome
Microbial gen
Phylogeny-evol-bioinf
Virus
Bioformatics/BLAST
Protein interact/gen syst
Plant gen
Basis: Cross-map of 50 c-clusters (bibliographic coupling) and 50 w-clusters (lexical coupling)
Method from Zitt, Lelu, Bassecoulard, Jasist 2011
Material from Laurens, Zitt, Bassecoulard, Scientometrics 2010

sources [2.257]. The detail of the more sophisticated techniques are not on the table: millions of Google users benefit from hybrid IR processes every day, but in spite of expansive literature devoted to the PageRank algorithm itself starting with [2.258] and published works on lexical/semantic processing [2.196], the detailed combination of multinetwork operations in the search engine is not documented. We will limit ourselves here to quite basic combinations, readily available in bibliometric literature.

The scope of hybridization is quite large: words and citations, on which we focus, may be taken either as variants of information tokens likely to be indistinctly treated under certain conditions, in a typical informetric posture; or seen as elements of quite different relations with their own fundamental properties and interpretation, suggesting their use in sequential or parallel protocols. Parallel exploitation, particularly, is *sociology-compatible* allowing for separate interpretations and comparison before final combination if necessary.

Full Hybrid

The structuring/clustering of fields using a common metric mixing citation and term distances at the finer grain level, from the start, is a promising path [2.259, 260]. *Boyack* and *Klavans* [2.128] on a large dataset, observed that even a *hybrid naive* coupling outperformed pure bibliographic coupling. Statistical differences between word and citation distribution can be reduced through a normalization of the similarity measures with different distributions ([2.261] with later simplification in [2.262]) achieving a full and flexible integration. *Koopman* et al. [2.263] established cluster similarities using a combination of tokens, for comparing clustering solutions based on direct vocabulary and indirect vocabulary associated with authors, journals, citation, etc.

Those developments remain within the framework of *feature methods* keeping the substance of information elements, words, and citations. In Sect. 2.3, we mentioned purely computational methods (character n-grams on text flow, compression) for calculating generalized text distances regardless of linguistic features. An option is to stay within the textual domain (full text, abstract, title...) or to enlarge to the full article including authors, affiliations, list of references, etc. We get a massive and blind form of hybridization, dissolving both terms and references into signals, ignoring all forms of normalization including zones length (text versus bibliography). Such black boxes are deprived of any semantic interpretation, but in our experience prove efficient for quick calculation of interdocument distances.

We have seen above (Sect. 2.3.2) that deep neural networks have proven in many areas of supervised learning, including information retrieval, their ability to do without prior weighting of the variables. Their unsupervised variants, building upon their success in very constrained fields like the Go game, should be able to do the same from an informal collection of data—such as *full hybrid* data—and so an application to domain delineation might be to consider the last layers of a network collecting the many traces of scientific activity: whatever citations, texts, and so on in the wake of present limited attempts of hybridization. Research in unsupervised deep learning, though, is still at a preliminary stage [2.264]. There is no doubt, however, that in the next few years progress—and controversy—are to be expected from deep learning's entry into the competition. These processes, however, remain black boxes, with quite difficult interpretations. Perhaps high-level semantic categorization resulting from the careful interpretation of the last layers might allow experts to select a subset of explicit dimensions in order to take into account the users' expectations of a delineation process. Whether this could reconcile cognitive classification and institutional expectations, an issue mentioned above, is another question.

Sequential Hybrid: Citations → Terms

Sequential protocols of delineation may rely on more iterations; we limit ourselves here to pointing out the basic sequences. We mentioned the tradition of completing citation objects by textual tagging above. The question of the validity of cocitation research fronts (Sect. 2.2.3) triggered further developments in terms of retrieval and recall rate and the means to foster it, possibly with the help of texts. *Braam* et al. [2.265] developed a systematic complementation of cocitation cluster coverage by lexical means, a first operational example of hybrid delineation. The citation → text sequence keeps being explored for other purposes, especially in global science maps. *Boyack* and *Klavans* [2.91] use textual metrics for display of cocitation cluster relations at the large scale where citation signals are weak.

Sequential Hybrid: Terms → Citations

Here, the perspective is reversed. The remote ancestor is a classical application of citation indexing, when title words or KeyWords Plus™ were used to query a citation index to harvest papers on a given (set of) topics. The rationale is simple: starting a multistep process with experts' help is easier with word queries. In a second step, the expansion is carried out on the citation network, where unsupervised or lightly supervised procedures are safer than on texts, with proper precautions.

General conditions for citation analysis are required, especially not too scarce reference lists. There is some analogy with the *boomerang effect* on citations [2.266]. An example of protocol is the Lex+Cite process explored in *Laurens* et al. [2.241], especially for emerging or transverse domains, where classical methods tend to fall short.

Quite a few options exist for expansion. If the seed is considered globally, literature with reference combinations present in the seed, but not in particular papers, is recalled. However, unspecific cites should be ruled out, which may require information from the whole database. Conversely, if only combinations at the paper level are allowed (strict bibliographic coupling), some broad-scope literature is missed; cluster-level enrichment, if a previous breakdown into clusters is available, stands in the middle. Besides the recall-oriented aim, these hybrid protocols may also enhance precision by submitting the core itself to bibliographic coupling constraints. Along the same lines, an elaborate strategy starting with lexical queries and query expansion, completed by journal selection and ending by collecting citing papers, is proposed in [2.267].

Parallel Design

As described above in Sect. 2.4.3, parallel design allows for comparison especially when metrics and clustering methods are identical, so that the final outcomes can be compared by factor analyses, parallel clustering–mapping, and reordered cross-tabulations. In parallel clustering, a similarity between clusters from different origins is defined after their degree of overlap, and then the intercluster matrix, of small size, is easily displayed using an MDS-type method. The cross-tabulation for example highlights strong relative overlaps with two strategies in addition to choosing either the c-cluster or the w-cluster on a topic: (a) precision-oriented: a heavy intersection between c-cluster and w-cluster suggests a strong form of topic, strategy possibly extended to superclusters (b) recall-oriented strategy, taking the union of c-cluster and w-cluster.

2.4.5 Conclusion

The various publication-linked networks, at least words and citations offer globally convergent views but not at the point that one can be happy with a single solution: sociology of citing, collaborating behavior and writing rhetorics keep some distance, and bibliometric protocols can choose to mix up all information tokens or to combine parallel approaches at the final stage only. Comparison and complementarity merit further endeavor. In practice, delineation cannot avoid supervision and actors' negotiation. Protocols of experts' guidance for evaluation purposes are desirable. Cross-validation of parallel processes, and even in some cases of sequential processes [2.241], may alleviate the burden of multistep external validation. There are strong indications that multinetwork methods improve recall and offer richer substance to expert/user discussions, but more benchmark studies against ground truth are needed.

2.5 Delineation Schemes and Conclusion

2.5.1 Delineation Schemes

IR Search First

A scheme of a bibliometric study asking for careful delineation may be as follows:

- For memory's sake, selection of the expert/peers panel, matching the expected variety of the domain.
- Supervised IR search on specialized journals and specific vocabulary, aiming at precision, building up the core of the domain. Alternatively, use of cited cores at the article or author level. The granularity is, typically, the document level. In favorable cases, some partial query formulas are found in the literature.
- Query expansion or bibliometric expansion with citations (the latter usually requiring lighter supervision). The query expansion is conducted globally or query by query. Optionally, a round of data analysis/clustering can suggest query rephrasing or complementing (Fig. 2.6a,b).
- Evaluation of outcomes especially on the borderline. In multilevel processes, the border region typically stands between the high-precision cores/seeds (or low-recall expanded set) and the high-recall expanded set. Circles of expansion with expected relevance indexes (example in Sect. 2.4) enlighten decision-making, again optionally supported by thematic clustering/mapping.

Clustering/Mapping First

Regional overset maps are expected to contain all the target, and the decisions on border regions are typically made at the cluster level. Granularity obviously matters: we cannot expect that any high-level clustering of global or superlocal science will directly produce

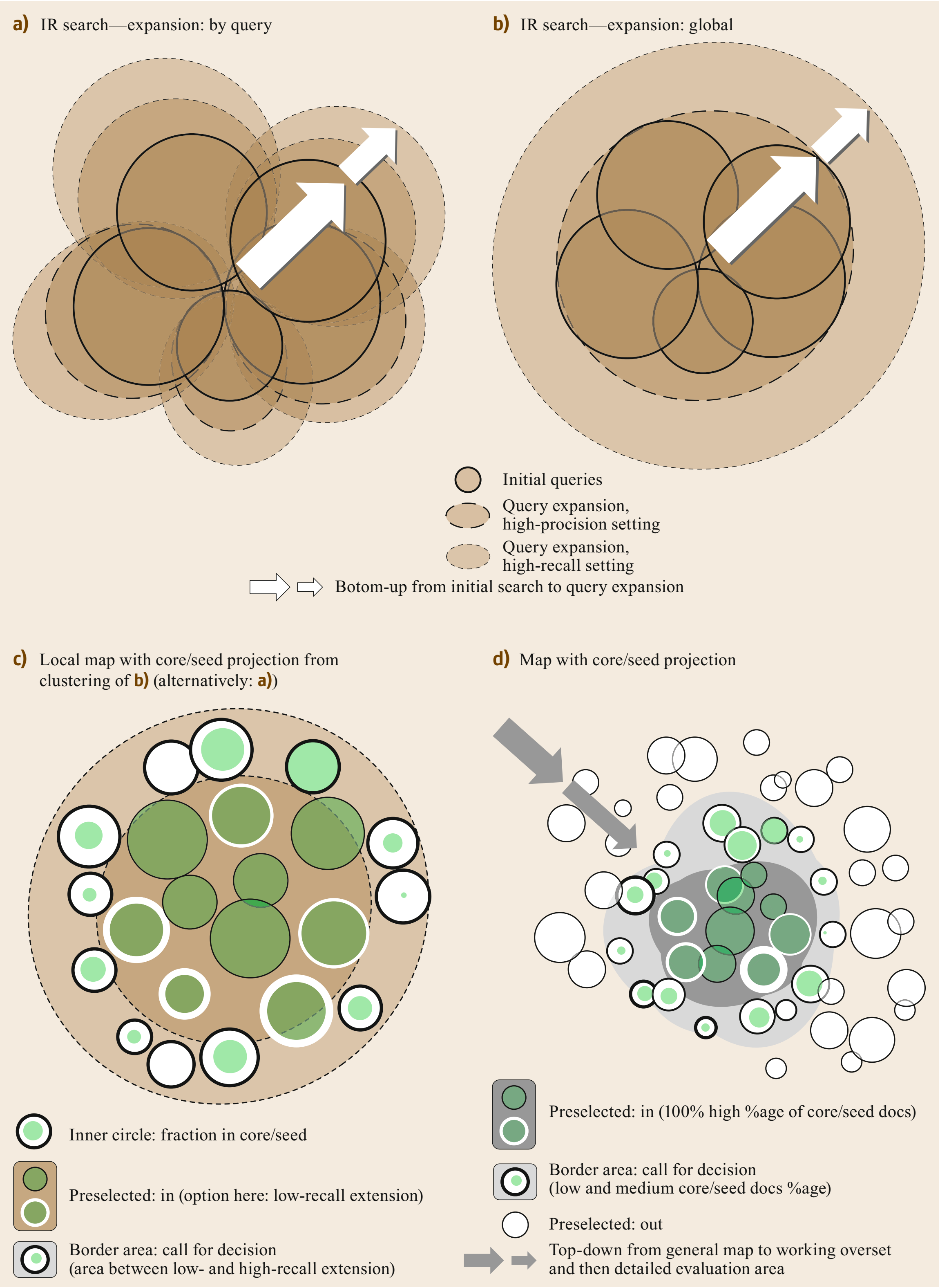

a) IR search—expansion: by query
b) IR search—expansion: global
Initial queries
Query expansion, high-procision setting
Query expansion, high-recall setting
Botom-up from initial search to query expansion
c) Local map with core/seed projection from clustering of b) (alternatively: a))
d) Map with core/seed projection
Inner circle: fraction in core/seed
Preselected: in (option here: low-recall extension)
Border area: call for decision (area between low- and high-recall extension)
Preselected: in (100% high %age of core/seed docs)
Border area: call for decision (low and medium core/seed docs %age)
Preselected: out
Top-down from general map to working overset and then detailed evaluation area

Fig. 2.6a–d IR search and mapping approaches. **(a)** IR process: bottom-up queries and expansion of individual queries. Assumed at the paper-level. **(b)** Variant of **(a)**: Expansion based on the entire set (lexical or citation-based). The border area, to be discussed, is typically determined by the region between the high-precision seed (or a low-recall expanded set) and a high-recall set. Circles of expansion in the border region, if indirect indicators or relevance are available, can drive the choice of delineation. (Optional **(c)**) Local clustering from B data (alternatively: from A data). Clusters are helpful for discussion but the border region and the decision tools may exist in A or B stages. The map is local and in the general case is not superimposable to a fraction of the global map D. A discussion on global versus local mapping is found in [2.179]. **(d)** Global mapping/clustering: top-down from global or overset map to the target; detection of border area with the help of information from projection. The example in the next panel (Fig. 2.6e) assumes no external information, it relies on clustering outcomes only to define core, periphery and outside regions ◄

a class retrieving the target domain as a whole. A lower-level breakdown yielding fine-grained delineation of the frontier will be preferred, with a number of subareas large enough to match the diversity of the domain and eventually increase precision, but small enough to make cluster-level expertise feasible. Reasonably, the granularity picked fulfils two objectives, aiding the delineation and preparing the study of the domain's subareas.

In the perspective of a cluster evaluation procedure, possibly time-consuming and costly, it is recommended that one relies on a lightly supervised preselection of the border region, located between the internal core, a priori deemed *in*, and the external zone deemed *out*. Depending on the clustering–mapping protocols chosen (see the sketch Fig. 2.6c,d), various solutions can address this preselection, for example:

- Clustering with IR search projection. For this preselection, most helpful is the simultaneous representation of a global map (or at least of an overset-map) obtained on one criterion and cluster-level properties on another criterion. The projection of local features over a large context is often used: in two-step protocols, seeds for example are projected on clusters in the expanded set [2.241] with the ratio of seed articles as the indicator for delineation. Another combination: a global map conveys a particular vision depending on the network represented and the methodological choices made, and the hits of an IR search on a lexical marker (with a generous setting for recall) alerts one to clusters of interest. In Fig. 2.2 for example, the central communities might be considered as belonging to a core, whereas distant colonies, on the borders, require evaluation. Such cases illustrate the complementarity of IR and mapping techniques for avoiding silence both on weak and strong signals, as mentioned above. An alleviated process uses the projection of specialized journal literature onto a global map [2.177]. Such processes help pinpoint clusters forming the border region as the *decision area* and/or suggest journals or groups of papers as candidates for extending a core. Clusters may also undergo a complementary stage of query expansion or bibliometric expansion, typically transforming—in a given universe—a hard partition into an overlapping structure. For the domain delineation, only the overlaps involving the border region will matter for the final outcome.
- Crossing methods. An alternative is the crossing of literature sets produced by different techniques or upon different networks. Instead of the standard core–periphery schemes, visualization may confront cognitive viewpoints, where areas of convergence (overlaps) are considered as strong forms (another form of core) and nonoverlapping parts as possible extensions to be validated. An example of crossmaps was shown in Sect. 2.4. In the limit case of Boolean formulas addressing the whole domain to delineate, this would be equivalent to running a word-based search AND/OR a citation-based search. The AND clause yields the strong form and the OR clause a possible expansion along two branches, words and citations.
  The principle can be extended in a pragmatic way, given that (a) data analysis methods are not very robust and tend to yield quite different outcomes; (b) data from different networks do not lead to identical results (polyrepresentation). Therefore the combination of methods, or the combination of networks, provides both ways to enhance precision (*strong forms* where outcomes of different reliable methods converge), and ways to enhance recall, in divergence areas, at some risk.
- Decision region and cluster evaluation (Fig. 2.6e):
  - Evaluation at the cluster level. Again, thematic clusters are understood here in a broad meaning, whatever the data analysis method used. As a rule, there is no ground truth making the evaluation of recall, precision, and F-scores or variants straightforward, so the relevance of each cluster has to be assessed by indirect indicators and/or supervision based on available cluster data. A light manual scrutiny can rely on cluster aggregate information such as label, pseudotitle recomposed from most specific

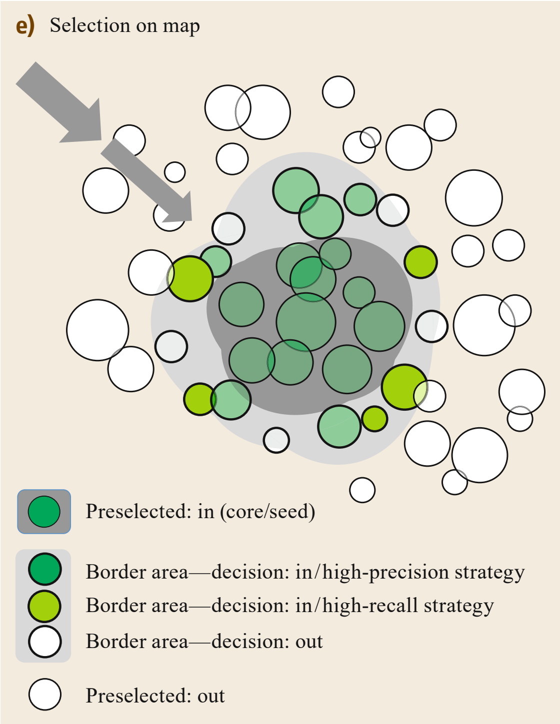


Fig. 2.6e IR search and mapping approaches. Evaluation and decision on clusters in border area. Example of direct selection of a bibliometric map on some criteria without input of other projected information

words or phrases, a ranked list of words, specific journals, cited authors/institutions, etc. Specificity of attributes is calculated by TF-IDF or other indexes. Features from a previous IR or mapping process, say ratios of expansion to core, or results from crossmaps, are particularly helpful. Map displays using pleasant interfaces make the task easier.

- Evaluation at a finer granularity level. Finer-grained information can be available from the delineation protocol: IR projections of good quality onto a map; cluster crossings from hybrid methods; combination with zones of bibliometric expansion, etc. In such cases the border region may be treated at the infracluster or the document level. In pure mapping exercises, the cluster level may simply reveal too coarse, with exceedingly large or heterogeneous groupings. In this case, one has to go deeper into cluster content, through sampling for detailed analysis or further breakdown, at a cost.

The driving of evaluation is conditioned by the mastering of methodological effects and biases, likely to yield very different outputs. A particular attention, at the domain level, should be brought to the tendency of metrics and methods to favor particular semantic dimensions: To what extent can a domain be extended towards its intellectual base, especially theoretical foundations? Towards its tools and techniques? Towards its objects and products? Decision rules, in absence of a IR standard, will be based on quantitative indicators of the process, for example the intensity of bibliometric linkages in expansion stages, and experts' advices in terms of subjective precision, recall, and their balance (tantamount to variants of F-score). The convergence of experts' preferences, with the help of self-rating, may be taken into account.

2.5.2 To Conclude

Delineation at the mesolevel deals with intermediary objects. Models in Price's tradition cast some light on the dynamics of the whole scientific system, whereas network theory proposes, at the microlevel, various mechanic models explaining emergence of mesostructures. The connection with practical solutions for topic and domain delineation, a rather multidisciplinary issue, will stimulate many research projects.

In practical studies, delineation operations should respect the proportionality principle. In simple cases, specialized and mature fields, the domain can be defined by using ready-made resources: official classifications, databases schemes. The complex cases which typically justify scientometric field studies—multidisciplinary, generic and emerging/unsettled domains—are precisely those where delineation and expertise are the more challenging. Coarse-grained approaches (journal-level) are easier to implement, but again hindered by a locally complex network and abundance of nonspecific media.

Bibliometrics both exploits and feeds science classification resources, literature searching and mapping models and human skill. Validation procedures include cross-analyses and direct supervision. The delineation tasks pull together multiple strands of bibliometrics and IR. They inherit progress in data and network analysis, as well as common limitations in data coverage, robustness issues, ergonomy challenges with respect to supervision and discussions with sponsors. Bibliometrics cannot pretend to operationalize in a standard manner all questions from decision-makers nor, in cognitive applications, all questions from sociologists of science and other scholars.

Within the scope where *bibliometric hypothesis* applies, a horizon of delineation is the comparison and combination of solutions from the networks which reflect scientific activity, essentially actors and institutions, citations and texts. Taking advantage of all available facets of data is a pragmatic choice, to which the concept of polyrepresentation has given a theoretical support. The cross-study of the three main universes associated to documents is also gaining attention in bibliometrics and sociology of research, supported by social network analysis. The theoretical profusion around models of growth and decline of communities is perhaps not settled now, but is very promising for understanding the invisible colleges in its various aspects. Will this multinetwork research track converge towards unified hypotheses? There is little doubt that progress in this matter will enlighten the delineation issue especially in emerging areas. Meanwhile, the question remains whether networks should be fully hybridized with more or less radical techniques—substantive or featureless—or various network solutions be conducted in parallel with final synthesis. In the background, the tremendous potential of deep learning on big science data is likely to reshuffle the cards in retrieval and classification methods. The prospects are unclear right now, as their lack of explainability is a serious drawback in the bibliometric delineation context.

The management of supervision is central to the feasibility of bibliometric studies and their delineation tasks. Configurations are diverse, one cannot compare simple problems requiring light supervision, with large studies on controversial areas. In the latter case, the operators of the study deal with a possibly complex managerial organization, with steering committees and expert panels mixing policy makers, stakeholders, and scientists, possibly with multiple roles. The selection of data sources and the methods of supervision, and finally the perimeter of the domain, will reflect those social stakes. The definition of fields or disciplines is particularly sensitive to academic interests, epistemic convictions and border issues, likely to create conflictual visions, sometimes between external observers and established players. The panel composition, to be efficient, should match the diversity of the domain, both in terms of thematic specialization and social stakes, with possibly some help from a few high-level generalists. In the mediation role, bibliometrics is also a social practice.

Bibliometric studies, if commissioned by administrations or institutions, enter a complex landscape of decision–help procedures where quantitative proposals are elements of discussion and decision among others. The question is vaster, however. *Gläser* et al. [2.26] underline the differences between operational definitions (say method outputs), pragmatic definitions (for clients and sponsors), and theoretical definitions (talking to science studies) of topics or domains. The notion of scientific domains is mobilized for a wide scope of purposes, labeling, information, and evaluation in scientific institutions, science administrations, IR databases of any kinds, laboratory life, scientists' self-positioning, and last but not least the reflexive work of scientometricians and social scientists on understanding the mechanisms of scientific activity.

References

2.1 A. Comte: *Cours de Philosophie Positive*, Vol. 1 (Rouen Frères, Paris 1830)
2.2 R.K. Merton: Science and technology in a democratic order, J. Leg. Political Sociol. **1**(1), 115–126 (1942)
2.3 R.K. Merton: *The Sociology of Science: Theoretical and Empirical Investigations* (Univ. Chicago Press, Chicago 1973)
2.4 T.S. Kuhn: *The Structure of Scientific Revolutions*, 2nd edn. (Univ. Chicago Press, Chicago 1970)
2.5 H.M. Collins, S. Yearley: Epistemological chicken. In: *Science as Practice and Culture*, ed. by A. Pickering (Univ. Chicago Press, Chicago 1992) pp. 301–326
2.6 B. Barnes, D. Bloor, J. Henry: *Scientific Knowledge: A Sociological Analysis* (Univ. Chicago Press, Chicago 1996)
2.7 D. Bloor: *Knowledge and Social Imagery* (Routledge Kegan Paul, London 1976)
2.8 K.D. Knorr-Cetina: Scientific communities or transepistemic arenas of research? A Critique of quasi-economic models of science, Soc. Stud. Sci. **12**(1), 101–130 (1982)
2.9 M.J. Mulkay, G.N. Gilbert, S. Woolgar: Problem areas and research networks in science, Sociology **9**(2), 187–203 (1975)
2.10 M. Serres: *La Traduction*, *Hermès III*, *Collection 'Critique'* (Les Éditions de Minuit, Paris 1974)
2.11 B. Latour, S. Woolgar: *Laboratory Life: The Social Construction of Scientific Facts* (SAGE, Beverly Hills 1979)
2.12 M. Callon, B. Latour: Unscrewing the big leviathan: How actors macro-structure reality and how sociologists help them to do so. In: *Advances in Social Theory and Methodology: Toward an Integration of Mirco- and Macro-Sociologies*, ed. by K. Knorr-Cetina, A.V. Cicourel (Routledge Kegan Paul, London 1981) pp. 277–303
2.13 J. Law, J. Hassard: *Actor Network Theory and After* (Blackwell, Oxford 1999)
2.14 T. Lenoir: *Instituting Science: The Cultural Production of Scientific Disciplines* (Stanford Univ. Press, Stanford 1997)

2.15 V. DiRita: Microbiology is an integrative field, so why are we a divided society?, Microbe Mag. **8**(10), 384–385 (2013)
2.16 A. Casadevall, F.C. Fang: Field science—The nature and utility of scientific fields, mBio **6**(5), e01259–15 (2015)
2.17 J. Piaget: L'épistémologie des relations interdisciplinaires. In: *Interdisciplinarity: Problems of Teaching and Research in Universities*, ed. by L. Apostel, G. Berger, A. Briggs, G. Michaud (OECD, Paris 1972) pp. 127–140
2.18 D.J.D. Price, D.D. Beaver: Collaboration in an invisible college, Am. Psychol. **21**(11), 1011–1018 (1966)
2.19 D. Crane: *Invisible Colleges: Diffusion of Knowledge in Scientific Communities* (Chicago Univ. Press, Chicago 1972)
2.20 D.E. Chubin: Beyond invisible colleges: Inspirations and aspirations of post-1972 social studies of science, Scientometrics **7**(3–6), 221–254 (1985)
2.21 A. Zuccala: Modeling the invisible college, J. Am. Soc. Inf. Sci. Technol. **57**(2), 152–168 (2005)
2.22 J. Gläser, G. Laudel: Integrating scientometric indicators into sociological studies: Methodical and methodological problems, Scientometrics **52**(3), 411–434 (2001)
2.23 P.M. Haas: Introduction: Epistemic communities and international policy coordination, Int. Organization **46**(1), 1–35 (1992)
2.24 É. Wenger: *Communities of Practice: Learning, Meaning, and Identity* (Cambridge Univ. Press, New York 1998)
2.25 R.P. Smiraglia: Domain analysis of domain analysis for knowledge organization: Observations on an emergent methodological cluster, Knowl. Organ. **42**(8), 602–611 (2015)
2.26 J. Gläser, A. Scharnhorst, W. Glänzel: Same data—Different results? Towards a comparative approach to the identification of thematic structures in science, Scientometrics **111**(2), 979–979 (2017)
2.27 C.R. Sugimoto, S. Weingart: The kaleidoscope of disciplinarity, J. Documentation **71**(4), 775–794 (2015)
2.28 R. Todorov: Representing a scientific field: A bibliometric approach, Scientometrics **15**(5/6), 593–605 (1989)
2.29 R.J.W. Tijssen: A quantitative assessment of interdisciplinary structures in science and technology: Co-classification analysis of energy research, Res. Policy **21**(1), 27–44 (1992)
2.30 C.S. Wagner: *The New Invisible College: Science for Development* (Brookings Institution, Washington 2008)
2.31 A. Suominen, H. Toivanen: Map of science with topic modeling: Comparison of unsupervised learning and human-assigned subject classification, J. Assoc. Inf. Sci. Technol. **67**(10), 2464–2476 (2016)
2.32 E.C.M. Noyons, A.F.J. van Raan: Monitoring scientific developments from a dynamic perspective: Self-organized structuring to map neural network research, J. Am. Soc. Inf. Sci. **49**(1), 68–81 (1998)
2.33 M. Zitt, E. Bassecoulard: Delineating complex scientific fields by an hybrid lexical-citation method: An application to nanosciences, Inf. Process. Manag. **42**(6), 1513–1531 (2006)
2.34 J.T. Klein: *Interdisciplinarity: History, Theory, and Practice* (Wayne State Univ. Press, Detroit 1990)
2.35 B.C.K. Choi, A.W.P. Pak: Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 1. Definitions, objectives, and evidence of effectiveness, Clin. Investig. Med. **29**(6), 351–364 (2006)
2.36 T. Jahn, M. Bergmann, F. Keil: Transdisciplinarity: Between mainstreaming and marginalization, Ecol. Econ. **79**, 1–10 (2012)
2.37 A.W. Russell, F. Wickson, A.L. Carew: Transdisciplinarity: Context, contradictions and capacity, Futures **40**(5), 460–472 (2008)
2.38 J.T. Klein: Evaluation of interdisciplinary and transdisciplinary research, Am. J. Prev. Med. **35**(2), S116–S123 (2008)
2.39 T.R. Miller, T.D. Baird, C.M. Littlefield, G. Kofinas, F.S. Chapin III, C.L. Redman: Epistemological pluralism: Reorganizing interdisciplinary research, Ecol. Soc. **13**(2), 46 (2008)
2.40 A. Yegros-Yegros, I. Rafols, P. D'Este: Does interdisciplinary research lead to higher citation impact? The different effect of proximal and distal interdisciplinarity, PLOS ONE **10**(8), e0135095 (2015)
2.41 G.E.A. Solomon, S. Carley, A.L. Porter: How multidisciplinary are the multidisciplinary journals science and nature?, PLOS ONE **11**(4), e0152637 (2016)
2.42 C.R. Sugimoto, N. Robinson-Garcia, R. Costas: Towards a global scientific brain: Indicators of researcher mobility using co-affiliation data. In: *OECD Blue Sky III Forum on Science and Innovation Indicators*, ed. by M. Feldman, S. Nagaoka, L. Soete, A. Jaffe, M. Salazar, R. Veugelers (OECD, Paris 2016)
2.43 M. Bordons, F. Morillo, I. Gómez: Analysis of cross-disciplinary research through bibliometric tools. In: *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems*, ed. by H.F. Moed, W. Glänzel, U. Schmoch (Springer, Dordrecht 2004) pp. 437–456
2.44 G. Pinski, F. Narin: Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics, Inf. Process. Manag. **12**(5), 297–312 (1976)
2.45 E.J. Rinia, T.N. van Leeuwen, E.E.W. Bruins, H.G. van Vuren, A.F.J. van Raan: Measuring knowledge transfer between fields of science, Scientometrics **54**(3), 347–362 (2002)
2.46 E. Bassecoulard, M. Zitt: Patents and publications: The lexical connection. In: *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems*, ed. by H.F. Moed, W. Glänzel, U. Schmoch (Springer, Dordrecht 2004) pp. 665–694
2.47 K. Börner, R. Klavans, M. Patek, A.M. Zoss, J.R. Biberstine, R.P. Light, V. Larivière, K.W. Boyack: Design and update of a classification system: The UCSD map of science, PLoS ONE **7**(7), e39464 (2012)

2.48 K.W. Boyack, R. Klavans: The structure of science. In: *Places and Spaces: Mapping Science—1st Iteration (2005): The Power of Maps*, ed. by K. Börner, D. MacPherson (scimaps.org, Indiana 2005)
2.49 A. Stirling: A general framework for analysing diversity in science, technology and society, J. R. Soc. Interface **4**(15), 707–719 (2007)
2.50 D. Hicks: Limitations and more limitations of co-citation analysis/bibliometric modelling: A reply to Franklin, Soc. Stud. Sci. **18**(2), 375–384 (1988)
2.51 H.F. Moed: *Citation Analysis in Research Evaluation, Information Science and Knowledge Management*, Vol. 9 (Springer, Dordrecht 2005)
2.52 A.F.J. van Raan, T.N. van Leeuwen, M.S. Visser: Severe language effect in university rankings: Particularly Germany and France are wronged in citation-based rankings, Scientometrics **88**(2), 495–498 (2011)
2.53 L. Soete, S. Schneegans, D. Eröcal, B. Angathevar, R. Rasiah: A world in search of an effective growth strategy. In: *UNESCO Science Report: Towards 2030, UNESCO Reference Works*, ed. by S. Schneegans (UNESCO, Paris 2015) pp. 20–55
2.54 J. Freyne, L. Coyle, B. Smyth, P. Cunningham: Relative status of journal and conference publications in computer science, Communications ACM **53**(11), 124–132 (2010)
2.55 A.J. Nederhof: Bibliometric monitoring of research performance in the social sciences and the humanities: A review, Scientometrics **66**(1), 81–100 (2006)
2.56 M. Huang, Y. Chang: Characteristics of research output in social sciences and humanities: From a research evaluation perspective, J. Am. Soc. Inf. Sci. Technol. **59**(11), 1819–1828 (2008)
2.57 G. Sivertsen, B. Larsen: Comprehensive bibliographic coverage of the social sciences and humanities in a citation index: An empirical analysis of the potential, Scientometrics **91**(2), 567–575 (2012)
2.58 T.N. Van Leeuwen, H.F. Moed, R.J.W. Tijssen, M.S. Visser, A.F.J. Van Raan: Language biases in the coverage of the Science Citation Index and its consequences for international comparisons of national research performance, Scientometrics **51**(1), 335–346 (2001)
2.59 M. Zitt, S. Ramanana-Rahary, E. Bassecoulard: Correcting glasses help fair comparisons in international science landscape: Country indicators as a function of ISI database delineation, Scientometrics **56**(2), 259–282 (2003)
2.60 V. Larivière, É. Archambault, Y. Gingras, É. Vignola-Gagné: The place of serials in referencing practices: Comparing natural sciences and engineering with social sciences and humanities, J. Am. Soc. Inf. Sci. Technol. **57**(8), 997–1004 (2006)
2.61 C. Michels, U. Schmoch: The growth of science and database coverage, Scientometrics **93**(3), 831–846 (2012)
2.62 S. Mikki: Comparing Google Scholar and ISI Web of Science for earth sciences, Scientometrics **82**(2), 321–331 (2010)
2.63 A. Sinha, Z.Y.S. Shen, H. Ma, D. Eide, B.J.P. Hsu, K. Wang: An overview of Microsoft Academic Service (MAS) and applications. In: *Proc. 24th Int. Conf. World Wide Web, Florence, Italy 2015*, ed. by A. Gangemi, S. Leonardi, A. Panconesi (ACM, New York 2015) pp. 243–246
2.64 D. Herrmannova, P. Knoth: An analysis of the Microsoft Academic Graph, D-Lib Mag. (2016), https://doi.org/10.1045/september2016-herrmannova
2.65 A.-W. Harzing, S. Alakangas: Microsoft academic: Is the phoenix getting wings?, Scientometrics **110**(1), 371–383 (2017)
2.66 J.E. Gray, M.C. Hamilton, A. Hauser, M.M. Janz, J.P. Peters, F. Taggert: Scholarish: Google Scholar and its value to the sciences, Issues Sci. Technol. Librarianship (2012), https://doi.org/10.5062/F4MK69T9
2.67 C. Labbé: Ike Antkare, one of the great stars in the scientific firmament, ISSI Newsletter **6**(2), 48–52 (2010)
2.68 P. Jacsó: Metadata mega mess in Google Scholar, Online Inf. Rev. **34**(1), 175–191 (2010)
2.69 A.-W. Harzing, S. Alakangas: Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison, Scientometrics **106**(2), 787–804 (2016)
2.70 Q. Wang, L. Waltman: Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus, J. Informetrics **10**(2), 347–364 (2016)
2.71 M. Thelwall, S. Haustein, V. Larivière, C.R. Sugimoto: Do altmetrics work? Twitter and ten other social web services, PLoS ONE **8**(5), e64841 (2013)
2.72 S. Haustein, I. Peters, J. Bar-Ilan, J. Priem, H. Shema, J. Terliesner: Coverage and adoption of altmetrics sources in the bibliometric community, Scientometrics **101**(2), 1145–1163 (2014)
2.73 E. Mohammadi, M. Thelwall: Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows, J. Assoc. Inf. Sci. Technol. **65**(8), 1627–1638 (2014)
2.74 Z. Zahedi, R. Costas, P. Wouters: How well developed are altmetrics? A cross-disciplinary analysis of the presence of “alternative metrics” in scientific publications, Scientometrics **101**(2), 1491–1513 (2014)
2.75 C.L. González-Valiente, J. Pacheco-Mendoza, R. Arencibia-Jorge: A review of altmetrics as an emerging discipline for research evaluation, Learn. Publ. **29**(4), 229–238 (2016)
2.76 A.E. Williams: Altmetrics: An overview and evaluation, Online Inf. Rev. **41**(3), 311–317 (2017)
2.77 C. Daraio, W. Glänzel: Grand challenges in data integration–state of the art and future perspectives: An introduction, Scientometrics **108**(1), 391–400 (2016)
2.78 OECD: *Revised Field of Science and Technology (FOS) Classification in the Frascati Manual—Report number DSTI/EAS/STP/NESTI(2006)19/FINAL* (OECD, Paris 2007)
2.79 E. Garfield: The evolution of the Science Citation Index, Int. Microbiol. **10**(1), 65–69 (2007)

2.80 A.I. Pudovkin, E. Garfield: Algorithmic procedure for finding semantically related journals, J. Am. Soc. Inf. Sci. Technol. **53**(13), 1113–1119 (2002)
2.81 E. Garfield: Citation analysis as a tool in journal evaluation: Journals can be ranked by frequency and impact of citations for science policy studies, Science **178**(4060), 471–479 (1972)
2.82 E. Garfield: The history and meaning of the journal impact factor, J. Am. Med. Assoc. **295**(1), 90–93 (2006)
2.83 F. Narin, G. Pinski, H.H. Gee: Structure of the biomedical literature, J. Am. Soc. Inf. Sci. **27**(1), 25–45 (1976)
2.84 P. Jacsó: As we may search: Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases, Curr. Sci. **89**(9), 1537–1547 (2005)
2.85 F. de Moya-Anegón, Z. Chinchilla-Rodríguez, B. Vargas-Quesada, E. Corera-Álvarez, F.J. Muñoz-Fernández, A. González-Molina, V. Herrero-Solana: Coverage analysis of Scopus: A journal metric approach, Scientometrics **73**(1), 53–78 (2007)
2.86 L. Leydesdorff, S.E. Cozzens: The delineation of specialties in terms of journals using the dynamic journal set of the SCI, Scientometrics **26**(1), 135–156 (1993)
2.87 E. Bassecoulard, M. Zitt: Indicators in a research institute: A multi-level classification of scientific journals, Scientometrics **44**(3), 323–345 (1999)
2.88 I. Rafols, M. Meyer: Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience, Scientometrics **82**(2), 263–287 (2009)
2.89 W. Glänzel, A. Schubert: A new classification scheme of science fields and subfields designed for scientometric evaluation purposes, Scientometrics **56**(3), 357–367 (2003)
2.90 E. Archambault, O.H. Beauchesne, J. Caruso: Towards a multilingual, comprehensive and open scientific journal ontology. In: *ISSI'11: Proc. 13th Int. Conf. Int. Soc. Scientometr. Informetrics, Durban, South Africa 2011*, ed. by E. Noyons, P. Ngulube, J. Leta (ISSI, Leiden Univ. Zululand 2011) pp. 66–77
2.91 K.W. Boyack, R. Klavans: Creation of a highly detailed, dynamic, global model and map of science, J. Assoc. Inf. Sci. Technol. **65**(4), 670–685 (2014)
2.92 R. Klavans, K.W. Boyack: Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge?, J. Assoc. Inf. Sci. Technol. **68**(4), 984–998 (2017)
2.93 A. Ruiz-Iniesta, O. Corcho: A review of ontologies for describing scholarly and scientific documents. In: *SePublica'14: Proc. 4th Workshop on Semantic Publishing Co-Located with the 11th Extended Semantic Web Conference, Anissaras, Greece 2014* (CEUR-WS, Aachen 2014), http://ceur-ws.org/Vol.1155/paper-07.pdf
2.94 A.M. Petersen, D. Rotolo, L. Leydesdorff: A triple helix model of medical innovation: Supply, demand, and technological capabilities in terms of medical subject headings, Res. Policy **45**(3), 666–681 (2016)
2.95 A. Mogoutov, B. Kahane: Data search strategy for science and technology emergence: A scalable and evolutionary query for nanotechnology tracking, Res. Policy **36**(6), 893–903 (2007)
2.96 A.L. Porter, J. Youtie, P. Shapira, D.J. Schoeneck: Refining search terms for nanotechnology, J. Nanoparticle Res. **10**(5), 715–728 (2007)
2.97 P. Ingwersen: Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory, J. Documentation **52**(1), 3–50 (1996)
2.98 J. Nicolaisen, B. Hjørland: Practical potentials of Bradford's law: A critical examination of the received view, J. Documentation **63**(3), 359–377 (2007)
2.99 P. Ingwersen, K. Järvelin: *The Turn: Integration of Information Seeking and Retrieval in Context*, The Information Retrieval Series, Vol. 18 (Springer, Dordrecht 2005)
2.100 T.E. Nisonger: Journals in the core collection: Definition, identification, and applications, Ser. Libr. **51**(3/4), 51–73 (2007)
2.101 H. Small: Co-citation in the scientific literature: A new measure of the relationship between two documents, J. Am. Soc. Inf. Sci. **24**(4), 265–269 (1973)
2.102 Q.L. Burrell: On the *h*-index, the size of the Hirsch core and Jin's *a*-index, J. Informetrics **1**(2), 170–177 (2007)
2.103 W. Glänzel, B. Thijs: Using "core documents" for detecting and labelling new emerging topics, Scientometrics **91**(2), 399–416 (2012)
2.104 J. Rocchio: Relevance feedback in information retrieval. In: *The SMART Retrieval System: Experiments in Automatic Document Processing*, ed. by G. Salton (Prentice Hall, Englewood Cliffs 1971) pp. 313–323
2.105 G. Salton, C. Buckley: Improving retrieval performance by relevance feedback, J. Am. Soc. Inf. Sci. **41**(4), 288–297 (1990)
2.106 C. Carpineto, G. Romano: A survey of automatic query expansion in information retrieval, ACM Comput. Surv. **44**(1), 1–50 (2012)
2.107 R. Agrawal, T. Imieliński, A. Swami: Mining association rules between sets of items in large databases, ACM SIGMOD Rec. **22**(2), 207–216 (1993)
2.108 D. Hric, R.K. Darst, S. Fortunato: Community detection in networks: Structural communities versus ground truth, Phys. Rev. E **90**(6), 062805 (2014)
2.109 M.M. Kessler: Bibliographic coupling between scientific papers, Am. Doc. **14**(1), 10–25 (1963)
2.110 N. Jardine, C.J. van Rijsbergen: The use of hierarchic clustering in information retrieval, Inf. Storage Retr. **7**(5), 217–240 (1971)
2.111 P. Mayr, A. Scharnhorst: Combining bibliometrics and information retrieval: Preface, Scientometrics **102**(3), 2191–2192 (2015)
2.112 P. Mayr, A. Scharnhorst: Scientometrics and information retrieval: Weak-links revitalized, Scientometrics **102**(3), 2193–2199 (2015)
2.113 M. Zitt: Meso-level retrieval: IR-bibliometrics interplay and hybrid citation-words methods in scientific fields delineation, Scientometrics **102**(3), 2223–2245 (2015)

2.114 P. Mayr, I. Frommholz, G. Cabanac, M.K. Chandrasekaran, K. Jaidka, M.-Y. Kan, D. Wolfram: Introduction to the special issue on bibliometric-enhanced information retrieval and natural language processing for digital libraries (BIRNDL), Int. J. Digit. Libr. **19**(2–3), 107–111 (2018)
2.115 M.E.J. Newman: The structure of scientific collaboration networks, Proc. Nat. Acad. Sci. **98**(2), 404–409 (2001)
2.116 M.E.J. Newman: Coauthorship networks and patterns of scientific collaboration, Proc. Nat. Acad. Sci. **101**(Suppl. 1), 5200–5205 (2004)
2.117 A.-L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, T. Vicsek: Evolution of the social network of scientific collaborations, Physica A: Stat. Mech. Appl. **311**(3/4), 590–614 (2002)
2.118 D.J.D. Price: A general theory of bibliometric and other cumulative advantage processes, J. Am. Soc. Inf. Sci. **27**(5), 292–306 (1976)
2.119 R. Albert, A.-L. Barabási: Statistical mechanics of complex networks, Rev. Mod. Phys. **74**(1), 47–97 (2002)
2.120 C.S. Wagner, L. Leydesdorff: Network structure, self-organization, and the growth of international collaboration in science, Res. Policy **34**(10), 1608–1618 (2005)
2.121 G. Csányi, B. Szendrői: Fractal-small-world dichotomy in real-world networks, Phys. Rev. E **70**(1), 016122 (2004)
2.122 M. McPherson, L. Smith-Lovin, J.M. Cook: Birds of a feather: Homophily in social networks, Annu. Rev. Sociol. **27**(1), 415–444 (2001)
2.123 N. Carayol, P. Roux: Knowledge flows and the geography of networks: A strategic model of small world formation, J. Econ. Behav. Organ. **71**(2), 414–427 (2009)
2.124 K. Börner, W. Glänzel, A. Scharnhorst, P. van den Besselaar: Modeling science: Studying the structure and dynamics of science, Scientometrics **89**(1), 347–348 (2011)
2.125 M. Cadot, A. Lelu, M. Zitt: Benchmarking 17 clustering methods, https://hal.archives-ouvertes.fr/hal-01532894 (2018)
2.126 A. McCallum, K. Nigam, L.H. Ungar: Efficient clustering of high-dimensional data sets with application to reference matching. In: *KDD'00: Proc. 6th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Boston, MA 2000*, ed. by R. Ramakrishnan, S. Stolfo, R. Bayardo, I. Parsa (Association for Computing Machinery, New York 2000) pp. 169–178
2.127 M. Zitt, E. Bassecoulard: Reassessment of co-citation methods for science indicators: Effect of methods improving recall rates, Scientometrics **37**(2), 223–244 (1996)
2.128 K.W. Boyack, R. Klavans: Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately?, J. Am. Soc. Inf. Sci. Technol. **61**(12), 2389–2404 (2010)
2.129 G.W. Milligan: A review of Monte Carlo tests of cluster analysis, Multivar. Behav. Res. **16**(3), 379–407 (1981)
2.130 G.W. Milligan, M.C. Cooper: Methodology review: Clustering methods, Appl. Psychol. Meas. **11**(4), 329–354 (1987)
2.131 M. Ester, H.-P. Kriegel, J. Sander, X. Xu: A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise. In: *KDD'96: Proc. 2nd Int. Con. Knowl. Discovery Data Mining, Portland, OR 1996*, ed. by E. Simoudis, J. Han, U. Fayyad (AAAI, Palo Alto 1996) pp. 226–231
2.132 A. Rodriguez, A. Laio: Clustering by fast search and find of density peaks, Science **344**(6191), 1492–1496 (2014)
2.133 M. Reinert: Un logiciel d'analyse lexicale, Cah. Ana. Données **11**(4), 471–481 (1986)
2.134 J.-P. Benzécri: *L'analyse des Correspondances, Analyse des Données*, Vol. 2 (Dunod, Paris 1973)
2.135 P.D. Turney, P. Pantel: From frequency to meaning: Vector space models of semantics, J. Artif. Intell. Res. **37**(1), 141–188 (2010)
2.136 S. Deerwester, S.T. Dumais, T.K. Landauer, G.W. Furnas, L. Beck: Improving information retrieval with latent semantic indexing. In: *Proc. 51st Annu. Meet. Am. Soc. Inf. Sci., Atlanta* (1988) pp. 36–40
2.137 A. Lelu: Clusters and factors: Neural algorithms for a novel representation of huge and highly multidimensional data sets. In: *New Approaches in Classification and Data Analysis*, ed. by E. Diday, Y. Lechevallier, M. Schader, P. Bertrand (Springer, Berlin 1994) pp. 241–248
2.138 C.H. Papadimitriou, G. Tamaki, P. Raghavan, S. Vempala: Latent semantic indexing: A probabilistic analysis. In: *PODS'98: Proc. 17th ACM SIGACT-SIGMOD-SIGART Symp. Principles Database Syst., Seattle, WA 1998*, ed. by A. Mendelson, J. Paredaens (ACM, New York 1998) pp. 159–168
2.139 T. Hofmann: Probabilistic latent semantic indexing. In: *SIGIR'99: Proc. 22nd Annu. Int. ACM SIGIR Conf. Res. Dev. Inf, Retrieval, Berkeley, CA 1999*, ed. by F. Gey, M. Hearst, R. Tong (ACM, New York 1999) pp. 50–57
2.140 D.M. Blei, A.Y. Ng, M.I. Jordan: Latent Dirichlet allocation, J. Mach. Learning Res. **3**, 993–1022 (2003)
2.141 V.D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre: Fast unfolding of communities in large networks, J. Stat. Mech. Theory Exp. **2008**(10), P10008 (2008)
2.142 M. Rosvall, C.T. Bergstrom: An information-theoretic framework for resolving community structure in complex networks, Proc. Natl. Acad. Sci. **104**(18), 7327–7331 (2007)
2.143 N.J. van Eck, L. Waltman: Software survey: VOSviewer, a computer program for bibliometric mapping, Scientometrics **84**(2), 523–538 (2010)
2.144 M. Meila, J. Shi: Learning segmentation by random walks. In: *NIPS'00: Proc. Neural Inf. Process. Syst. Conf., Denver, CO 2000*, ed. by T.K. Leen, T.G. Dietterich, V. Tresp (MIT Press, Cambridge 2000) pp. 873–879
2.145 A. Lancichinetti, S. Fortunato: Community detection algorithms: A comparative analysis, Phys. Rev. E **80**(5), 056117 (2009)

2.146 J. Leskovec, K.J. Lang, M. Mahoney: Empirical comparison of algorithms for network community detection. In: *WWW'10: Proc. 19th Int. Conf. World Wide Web, Raleigh, NC 2010*, ed. by M. Rappa, P. Jones, J. Freire, S. Chakrabarti (ACM, New York 2010) pp. 631–640

2.147 J. Yang, J. Leskovec: Defining and evaluating network communities based on ground-truth. In: *ICDM'12: Proc. 12th Int. Conf. Data Mining, Brussels 2012*, ed. by M.J. Zaki, A. Siebes, J.X. Yu, B. Goethals, G. Webb, X. Wu (IEEE, Los Alamitos 2012) pp. 745–754

2.148 Y. Shen, X. He, J. Gao, L. Deng, G. Mesnil: A latent semantic model with convolutional-pooling structure for information retrieval. In: *CIKM'14: Proc. 23rd ACM Conf. Inf. Knowl. Mining, Shanghai 2014*, ed. by J. Li, X.S. Wang, M. Garofalakis, I. Soboroff, T. Suel, M. Wang (ACM, New York 2014) pp. 101–110

2.149 C. Van Gysel, M. de Rijke, E. Kanoulas: Neural vector spaces for unsupervised information retrieval, ACM Trans. Inf. Syst. **36**(4), 1–25 (2018)

2.150 T. Mikolov, W. tau Yih, G. Zweig: Linguistic regularities in continuous space word representations. In: *NAACL-HLT'13: Proc. Conf. North Am. Chap. Assoc. Comput. Linguistics: Human Lang. Technol., Atlanta, GA 2013*, ed. by L. Vanderwende, H. Daume III, K. Kirchhoff (Association for Computational Linguistics, Stroudsburg 2013) pp. 746–751

2.151 O. Levy, Y. Goldberg: Neural word embedding as implicit matrix factorization. In: *NIPS'14: Proc. Neural Inf. Process. Syst. Conf., Monreéal 2014*, ed. by Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence (Curran Associates, Red Hook 2014) pp. 2177–2185

2.152 S.E. Robertson, S. Walker, S. Jones, M. Hancock-Beaulieu, M. Gatford: Okapi at TREC-3. In: *TREC'94: Proc. 3rd Text Retrieval Conf., Gaithersburg, MA 1994*, ed. by D.K. Harman (NIST, Gaithersburg 1994) pp. 109–126

2.153 T.M.J. Fruchterman, E.M. Reingold: Graph drawing by force-directed placement, Softw. Pract. Exp. **21**(11), 1129–1164 (1991)

2.154 M. Bastian, S. Heymann, M. Jacomy: Gephi: An open source software for exploring and manipulating networks. In: *ICWSM'09: Proc. 3rd Int. AAAI Conf. Weblogs Soc. Media, San Jose, CA 2009*, ed. by W.W. Cohen, N. Nicolov (AAAI, Palo Alto 2009) p. 361

2.155 S. Martin, W.M. Brown, R. Klavans, K.W. Boyack: OpenOrd: An open-source toolbox for large graph layout. In: *Proc. Visualization Data Analysis 2011, San Francisco, CA 2011*, ed. by P.C. Wong, J. Park, M.C. Hao, C. Chen, K. Börner, D.L. Kao, J.C. Roberts (SPIE, Bellingham 2011) p. 786806

2.156 W. de Nooy, A. Mrvar, V. Batagelj: *Exploratory Social Network Analysis with Pajek, Revised and Expanded*, 2nd edn. (Cambridge Univ. Press, New York 2011)

2.157 M. Cadot, A. Lelu: Optimized representation for classifying qualitative data. In: *DBKDA'10: Proc. 2nd Int. Conf. Adv. Databases, Knowl., Data Applications, Les Menuires, France 2010*, ed. by F. Laux, L. Strömbäck (IEEE, Los Alamitos 2010) pp. 241–246

2.158 D. Cai, X. He, J. Han: Document clustering using locality preserving indexing, IEEE Trans. Knowl. Data Eng. **17**(12), 1624–1637 (2005)

2.159 W.M. Rand: Objective criteria for the evaluation of clustering methods, J. Am. Stat. Assoc. **66**(336), 846–850 (1971)

2.160 T.M. Cover, J.A. Thomas: *Elements of Information Theory*, Wiley Series in Telecommunications (Wiley, New York 1991)

2.161 P. Ronhovde, Z. Nussinov: Multiresolution community detection for megascale networks by information-based replica correlations, Phys. Rev. E **80**(1), 016109 (2009)

2.162 E. Garfield, A.I. Pudovkin, V.S. Istomin: Why do we need algorithmic historiography?, J. Am. Soc. Inf. Sci. Technol. **54**(5), 400–412 (2003)

2.163 I. Marshakova: System of document connections based on references, Nauchno-Tekh. Inf. 2 **6**, 3–8 (1973)

2.164 H.D. White, B.C. Griffith: Author cocitation: A literature measure of intellectual structure, J. Am. Soc. Inf. Sci. **32**(3), 163–171 (1981)

2.165 G. Salton: *The SMART Retrieval System: Experiments in Automatic Document Processing* (Prentice Hall, Englewood Cliffs 1971)

2.166 M. Callon, J.-P. Courtial, W.A. Turner, S. Bauin: From translations to problematic networks: An introduction to co-word analysis, Soc. Sci. Inf. **22**(2), 191–235 (1983)

2.167 W.A. Turner, G. Chartron, F. Laville, B. Michelet: Packaging information for peer review: New co-word analysis techniques. In: *Handbook of Quantitative Science and Technology*, ed. by A.F.J. van Raan (Springer, Dordrecht 1988) pp. 291–323

2.168 J. Whittaker: Creativity and conformity in science: Titles, keywords and co-word analysis, Soc. Stud. Sci. **19**(3), 473–496 (1989)

2.169 L.C. Freeman: *The Development of Social Network Analysis: A Study in the Sociology of Science* (Empirical, Vancouver 2004)

2.170 C. Chen: CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature, J. Am. Soc. Inf. Sci. Technol. **57**(3), 359–377 (2006)

2.171 W. Glänzel, H.-J. Czerwon: A new methodological approach to bibliographic coupling and its application to the national, regional and institutional level, Scientometrics **37**(2), 195–221 (1996)

2.172 L. Waltman, N.J. van Eck: A new methodology for constructing a publication-level classification system of science, J. Am. Soc. Inf. Sci. Technol. **63**(12), 2378–2392 (2012)

2.173 N. Shibata, Y. Kajikawa, Y. Takeda, K. Matsushima: Comparative study on methods of detecting research fronts using different types of citation, J. Am. Soc. Inf. Sci. Technol. **60**(3), 571–580 (2009)

2.174 B. Jarneving: A comparison of two bibliometric methods for mapping of the research front, Scientometrics **65**(2), 245–263 (2005)

2.175 K. Börner: *Atlas of Science: Visualizing What We Know* (MIT Press, Cambridge 2010)

2.176 M. Zitt, E. Bassecoulard: Development of a method for detection and trend analysis of research fronts built by lexical or cocitation analysis, Scientometrics **30**(1), 333–351 (1994)
2.177 L. Leydesdorff, I. Rafols: Interactive overlays: A new method for generating global journal maps from web-of-science data, J. Informetrics **6**(2), 318–332 (2012)
2.178 L. Leydesdorff, P. Zhou: Nanotechnology as a field of science: Its delineation in terms of journals and patents, Scientometrics **70**(3), 693–713 (2007)
2.179 K.W. Boyack: Investigating the effect of global data on topic detection, Scientometrics **111**(2), 999–1015 (2017)
2.180 C. Bergstrom: Eigenfactor: Measuring the value and prestige of scholarly journals, College Res. Libr. News **68**(5), 314–316 (2007)
2.181 M. Zitt, H. Small: Modifying the journal impact factor by fractional citation weighting: The audience factor, J. Am. Soc. Inf. Sci. Technol. **59**(11), 1856–1860 (2008)
2.182 L. Waltman, N.J. van Eck, T.N. van Leeuwen, M.S. Visser: Some modifications to the SNIP journal impact indicator, J. Informatrics **7**(2), 272–285 (2013)
2.183 M. Zitt, J.-P. Cointet: Citation impacts revisited: How novel impact measures reflect interdisciplinarity and structural change at the local and global level. In: *ISSI'13: Proc. 14th Int. Conf. Int. Soc. Scientometr. Informetrics, Vienna, Austria 2013*, ed. by J. Gorraiz, E. Schiebel (Austrian Institute of Technology, Vienna 2013) pp. 285–299
2.184 H. Small, E. Sweeney: Clustering the Science Citation Index® using co-citations: I. A comparison of methods, Scientometrics **7**(3–6), 391–409 (1985)
2.185 T. Luukkonen, R.J.W. Tijssen, O. Persson, G. Sivertsen: The measurement of international scientific collaboration, Scientometrics **28**(1), 15–36 (1993)
2.186 M. Zitt, E. Bassecoulard, Y. Okubo: Shadows of the past in international cooperation: Collaboration profiles of the top five producers of science, Scientometrics **47**(3), 627–657 (2000)
2.187 K.W. Boyack, R. Klavans, K. Börner: Mapping the backbone of science, Scientometrics **64**(3), 351–374 (2005)
2.188 G. Lewison, G. Paraje: The classification of biomedical journals by research level, Scientometrics **60**(2), 145–157 (2004)
2.189 S. Teufel, J. Carletta, M. Moens: An annotation scheme for discourse-level argumentation in research articles. In: *EACL'99: Proc. 9th Conf. Eur. Chap. Assoc. Comput. Linguistics, Bergen, Norway 1999*, ed. by H.S. Thompson, A. Lascarides (ACL, Stroudsburg 1999) pp. 110–117
2.190 M. Liakata, S. Saha, S. Dobnik, C. Batchelor, D. Rebholz-Schuhmann: Automatic recognition of conceptualization zones in scientific articles and two life science applications, Bioinformatics **28**(7), 991–1000 (2012)
2.191 S. Teufel, M. Moens: Summarizing scientific articles: Experiments with relevance and rhetorical status, Comput. Linguist. **28**(4), 409–445 (2002)
2.192 C. Lyon, J. Malcolm, B. Dickerson: Detecting short passages of similar text in large document collections. In: *EMNLP'01: Proc. Conf. Empirical Methods Natur. Lang. Process., Pittsburgh, PA 2001*, ed. by L. Lee, D. Harman (ACL, Stroudsburg 2001) pp. 118–125
2.193 R. Cilibrasi, P.M.B. Vitányi: Clustering by compression, IEEE Trans. Inf. Theory **51**(4), 1523–1545 (2005)
2.194 C.H. Bennett, P. Gács, M. Li, P.M.B. Vitányi, W.H. Zurek: Information distance, IEEE Trans. Inf. Theory **44**(4), 1407–1423 (1998)
2.195 M. Li, X. Chen, X. Li, B. Ma, P.M.B. Vitányi: The similarity metric, IEEE Trans. Inf. Theory **50**(12), 3250–3264 (2004)
2.196 R. Cilibrasi, P.M.B. Vitányi: The Google similarity distance, IEEE Trans. Knowl. Data Eng. **19**(3), 370–383 (2007)
2.197 J. MacQueen: Some methods for classification and analysis of multivariate observations. In: *Proc. 5th Berkeley Symp. Math. Statistics Probabilities, Durban, South Africa 1967*, ed. by L.M. Le Cam, J. Neyman (Univ. California, Berkeley 1967) pp. 281–297
2.198 E.W. Forgy: Cluster analysis of multivariate data: Efficiency versus interpretability of classifications, Biometrics **21**, 768–769 (1965)
2.199 N. Sahoo, J. Callan, R. Krishnan, G. Duncan, R. Padman: Incremental hierarchical clustering of text documents. In: *CIKM'06: Proc. 15th ACM Int. Conf. Inf. Knowl. Manag., Arlington, VA 2006*, ed. by P.S. Yu, V. Tsotras, E. Fox, B. Liu (ACM, New York 2006) pp. 357–366
2.200 H. Yu, D. Searsmith, X. Li, J. Han: Scalable construction of topic directory with nonparametric closed termset mining. In: *ICDM'04: Proc. 4th IEEE Int. Conf. Data Mining, Brighton 2004*, ed. by R. Rastogi, K. Morik, M. Bramer, X. Wu (IEEE, Los Alamitos 2004) pp. 1–4
2.201 F. Åström: Changes in the LIS research front: Time-sliced cocitation analyses of LIS journal articles, 1990–2004, J. Am. Soc. Inf. Sci. Technol. **58**(7), 947–957 (2007)
2.202 D.M. Blei, J.D. Lafferty: Dynamic topic models. In: *ICML'06: Proc. 23rd Int. Conf. Mach. Learning, Pittsburgh, PA 2006*, ed. by W.W. Cohen, A. Moore (ACM, New York 2006) pp. 113–120
2.203 Q. Mei, C.X. Zhai: Discovering evolutionary theme patterns from text: An exploration of temporal text mining. In: *KDD'05: Proc. 11th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining, Chicago, IL 2005*, ed. by R.L. Grossman, R. Bayardo, K. Bennett, J. Vaidya (Association for Computing Machinery, New York 2005) pp. 198–207
2.204 F. Janssens, W. Glänzel, B. De Moor: Dynamic hybrid clustering of bioinformatics by incorporating text mining and citation analysis. In: *KDD'07: Proc. 13th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining, San Jose, CA 2007*, ed. by P. Berkhin, R. Caruana, X. Wu, S. Gaffney (Association for Computing Machinery, New York 2007) pp. 360–369
2.205 C. Chen, F. Ibekwe-SanJuan, J. Hou: The structure and dynamics of cocitation clusters: A multiple-

perspective cocitation analysis, J. Am. Soc. Inf. Sci. Technol. **61**(7), 1386–1409 (2010)
2.206 D. Chavalarias, J.-P. Cointet: Phylomemetic patterns in science evolution—the rise and fall of scientific fields, PLOS ONE **8**(2), e54847 (2013)
2.207 S. Shane: Technological Opportunities and New Firm Creation, Manag. Sci. **47**(2), 205–220 (2001)
2.208 K.B. Dahlin, D.M. Behrens: When is an invention really radical? Defining and measuring technological radicalness, Res. Policy **34**(5), 717–737 (2005)
2.209 H. Small, H. Tseng, M. Patek: Discovering discoveries: Identifying biomedical discoveries using citation contexts, J. Informetrics **11**(1), 46–62 (2017)
2.210 E. Garfield, I.H. Sher: KeyWords Plus™—algorithmic derivative indexing, J. Am. Soc. Inf. Sci. **44**(5), 298–299 (1993)
2.211 R.N. Kostoff, J.A. del Río, J.A. Humenik, E.O. García, A.M. Ramírez: Citation mining: Integrating text mining and bibliometrics for research user profiling, J. Am. Soc. Inf. Sci. Technol. **52**(13), 1148–1156 (2001)
2.212 B. Verspagen, C. Werker: The invisible college of the economics of innovation and technological change, Estudios de Economía Aplicada **21**(3), 187–203 (1975)
2.213 D. D. Beaver, R. Rosen: Studies in scientific collaboration – Part I. The professional origins of scientific co-authorship, Scientometrics **1**(3), 231–245 (1979)
2.214 T. Luukkonen, O. Persson, G. Sivertsen: Understanding patterns of international scientific collaboration, Sci. Technol. Hum. Values **17**(1), 101–126 (1992)
2.215 H. Kretschmer: Coauthorship networks of invisible colleges and institutionalized communities, Scientometrics **30**(1), 363–369 (1994)
2.216 J.S. Katz, B.R. Martin: What is research collaboration?, Res. Policy **26**(1), 1–18 (1997)
2.217 J.S. Katz: Geographical proximity and scientific collaboration, Scientometrics **31**(1), 31–43 (1994)
2.218 J. Hoekman, K. Frenken, R.J.W. Tijssen: Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe, Res. Policy **39**(5), 662–673 (2010)
2.219 T. Velden, A. Haque, C. Lagoze: A new approach to analyzing patterns of collaboration in co-authorship networks: Mesoscopic analysis and interpretation, Scientometrics **85**(1), 219–242 (2010)
2.220 P. Mutschke, A.Q. Haase: Collaboration and cognitive structures in social science research fields. Towards socio-cognitive analysis in information systems, Scientometrics **52**(3), 487–502 (2001)
2.221 J. Raffo, S. Lhuillery: How to play the "Names Game": Patent retrieval comparing different heuristics, Res. Policy **38**(10), 1617–1627 (2009)
2.222 K.W. McCain: The author cocitation structure of macroeconomics, Scientometrics **5**(5), 277–289 (1983)
2.223 G.N. Gilbert: Referencing as Persuasion, Soc. Stud. Sci. **7**(1), 113–122 (1977)
2.224 C. Roth, J.-P. Cointet: Social and semantic coevolution in knowledge networks, Soc. Netw. **32**(1), 16–29 (2010)
2.225 X. Polanco, L. Grivel, J. Royauté: How to do things with terms in informetrics: Terminological variation and stabilization as science watch indicators. In: *ISSI'95: Proc. 5th Int. Conf. Int. Soc. Scientometr. Informetrics, River Forest, IL 1995*, ed. by M.E.D. Koenig, A. Bookstein (Learned Information, Medford 1995) pp. 435–444
2.226 M.F. Porter: An algorithm for suffix stripping, Program **14**(3), 130–137 (1980)
2.227 L. Egghe, R. Rousseau: *Introduction to Informetrics: Quantitative Methods in Library, Documentation, and Information Science* (Elsevier, Amsterdam 1990)
2.228 K.W. McCain: Descriptor and citation retrieval in the medical behavioral sciences literature: Retrieval overlaps and novelty distribution, J. Am. Soc. Inf. Sci. **40**(2), 110–114 (1989)
2.229 M.L. Pao: Term and citation retrieval: A field study, Inf. Process. Manag. **29**(1), 95–112 (1993)
2.230 L. Bornmann, H.-D. Daniel: What do citation counts measure? A review of studies on citing behavior, J. Documentation **64**(1), 45–80 (2008)
2.231 B. Cronin: *The Citation Process: The Role Significance of Citations in Scientific Communication* (Taylor Graham, London 1984)
2.232 H.G. Small: Cited documents as concept symbols, Soc. Stud. Sci. **8**(3), 327–340 (1978)
2.233 B. Latour: *Science in Action: How to Follow Scientists and Engineers Through Society* (Harvard Univ. Press, Cambridge 1987)
2.234 A. Cambrosio, P. Keating, S. Mercier, G. Lewison, A. Mogoutov: Mapping the emergence and development of translational cancer research, Eur. J. Cancer **42**(18), 3140–3148 (2006)
2.235 F. Narin, E. Noma: Is technology becoming science?, Scientometrics **7**(3–6), 369–381 (1985)
2.236 M. Callon: Pinpointing industrial invention: An exploration of quantitative methods for the analysis of patents. In: *Mapping the Dynamics of Science and Technology*, ed. by M. Callon, J. Law, A. Rip (Macmillan, Houndmills, London 1986) pp. 163–188
2.237 V. Lariviére, É. Archambault, Y. Gingras: Long-term variations in the aging of scientific literature: From exponential growth to steady-state science (1900–2004), J. Am. Soc. Inf. Sci. Technol. **59**(2), 288–296 (2008)
2.238 E.C.M. Noyons, R.K. Buter, A.F.J. van Raan, H. Schwechheimer, M. Winterhager, P. Weingart: *The Role of Europe in World-Wide Science and Technology: Monitoring and Evaluation in a Context of Global Competition—Report for the European Commission* (CWTS-Leiden, IWT-Bielefeld, Brussels 2000)
2.239 E.C.M. Noyons, R.K. Buter, A.F.J. van Raan, U. Schmoch, T. Heinze, S. Hinze, R. Rangnow: *Mapping Excellence in Science and Technology Across Europe Nanoscience and Nanotechnology—Report of Project EC-PPN CT-2002-0001 to the European Commission* (CWTS and Fraunhofer ISI, Leiden, Karlsruhe 2003)
2.240 M. Zitt, A. Lelu, E. Bassecoulard: Hybrid citation-word representations in science mapping: Portolan

charts of research fields?, J. Am. Soc. Inf. Sci. Technol. **62**(1), 19–39 (2011)
2.241 P. Laurens, M. Zitt, E. Bassecoulard: Delineation of the genomics field by hybrid citation-lexical methods: Interaction with experts and validation process, Scientometrics **82**(3), 647–662 (2010)
2.242 S.A. Morris, G.G. Yen: Crossmaps: Visualization of overlapping relationships in collections of journal papers, Proc. Natl. Acad. Sci. **101**(Suppl. 1), 5291–5296 (2004)
2.243 C. Reilly, C. Wang, M. Rutherford: A rapid method for the comparison of cluster analyses, Stat. Sin. **15**(1), 19–33 (2005)
2.244 R. Klavans, K.W. Boyack: Toward a consensus map of science, J. Am. Soc. Inf. Sci. Technol. **60**(3), 455–476 (2009)
2.245 L. Leydesdorff, I. Rafols: A global map of science based on the ISI subject categories, J. Am. Soc. Inf. Sci. Technol. **60**(2), 348–362 (2009)
2.246 P. Ahlgren, B. Jarneving: Bibliographic coupling, common abstract stems and clustering: A comparison of two document-document similarity approaches in the context of science mapping, Scientometrics **76**(2), 273–290 (2008)
2.247 E. Yan, Y. Ding: Scholarly network similarities: How bibliographic coupling networks, citation networks, cocitation networks, topical networks, coauthorship networks, and coword networks relate to each other, J. Am. Soc. Inf. Sci. Technol. **63**(7), 1313–1326 (2012)
2.248 T. Velden, K.W. Boyack, J. Gläser, R. Koopman, A. Scharnhorst, S. Wang: Comparison of topic extraction approaches and their results, Scientometrics **111**(2), 1169–1221 (2017)
2.249 H. Small: Co-Citation Context Analyses and the Structure of Paradigms, J. Documentation **36**(3), 183–196 (1980)
2.250 H. Small: Maps of science as interdisciplinary discourse: Co-citation contexts and the role of analogy, Scientometrics **83**(3), 835–849 (2010)
2.251 S. Teufel, A. Siddharthan, D. Tidhar: Automatic classification of citation function. In: *EMNLP'06: Proc. Conf. Empirical Methods Nat. Lang. Process., Sydney, Australia 2006*, ed. by D. Jurafsky, É. Gaussier (ACL, Stroudsburg 2006) pp. 103–110
2.252 A. Ritchie, S. Robertson, S. Teufel: Comparing citation contexts for information retrieval. In: *CIKM'08: Proc. 17th ACM Conf. Inf. Knowl. Mining, Napa Valley, CA 2008*, ed. by J.G. Shanahan, S. A.-Yahia, I. Manolescu, Y. Zhang, D.A. Evans, A. Kolcz, K.-S. Choi, A. Chowdury (Association for Computing Machinery, New York 2008) pp. 213–222
2.253 S. Liu, C. Chen: The differences between latent topics in abstracts and citation contexts of citing papers, J. Am. Soc. Inf. Sci. Technol. **64**(3), 627–639 (2013)
2.254 H. Small: Interpreting maps of science using citation context sentiments: A preliminary investigation, Scientometrics **87**(2), 373–388 (2011)
2.255 A. Elkiss, S. Shen, A. Fader, G. Erkan, D. States, D. Radev: Blind men and elephants: What do citation summaries tell us about a research article?, J. Am. Soc. Inf. Sci. Technol. **59**(1), 51–62 (2008)
2.256 A. Callahan, S. Hockema, G. Eysenbach: Contextual cocitation: Augmenting cocitation analysis and its applications, J. Am. Soc. Inf. Sci. Technol. **61**(6), 1130–1143 (2010)
2.257 X. He, C.H.Q. Ding, H. Zha, H.D. Simon: Automatic topic identification using webpage clustering. In: *ICDM'01: Proc. Int. Conf. Data Mining, San Jose, CA 2001*, ed. by N. Cercone, T.Y. Lin, X. Wu (IEEE, Los Alamitos 2001) pp. 195–202
2.258 S. Brin, L. Page: The anatomy of a large-scale hypertextual web search engine, Comp. Netw. ISDN Syst. **30**(1–7), 107–117 (1998)
2.259 P. van den Besselaar, G. Heimeriks: Mapping research topics using word-reference co-occurrences: A method and an exploratory case study, Scientometrics **68**(3), 377–393 (2006)
2.260 P. Ahlgren, C. Colliander: Document–document similarity approaches and science mapping: Experimental comparison of five approaches, J. Informetrics **3**(1), 49–63 (2009)
2.261 F. Janssens, W. Glänzel, B. De Moor: A hybrid mapping of information science, Scientometrics **75**(3), 607–631 (2008)
2.262 W. Glänzel, B. Thijs: Using "core documents" for the representation of clusters and topics, Scientometrics **88**(1), 297–309 (2011)
2.263 R. Koopman, S. Wang, A. Scharnhorst: Contextualization of topics: Browsing through the universe of bibliographic information, Scientometrics **111**(2), 1119–1139 (2017)
2.264 Y. LeCun: A path to AI. In: *BAI'17: Workshop Beneficial Artif. Intell., Asilomar, CA 2017*, ed. by E. Brynjolfsson, E. Horvitz, P. Norvig, F. Rossi, S. Russell, B. Selman (Future of Life Institute, Cambridge 2017), https://futureoflife.org/wp-content/uploads/2017/01/Yann-LeCun.pdf
2.265 R.R. Braam, H.F. Moed, A.F.J. van Raan: Mapping of science by combined co-citation and word analysis. I. Structural aspects, J. Am. Soc. Inf. Sci. **42**(4), 233–251 (1991)
2.266 B. Larsen: Exploiting citation overlaps for information retrieval: Generating a boomerang effect from the network of scientific papers, Scientometrics **54**(2), 155–178 (2002)
2.267 Y. Huang, J. Schuehle, A.L. Porter, J. Youtie: A systematic method to create search strategies for emerging technologies based on the Web of Science: Illustrated for "Big Data", Scientometrics **105**(3), 2005–2022 (2015)

Michel Zitt

Lereco Lab (retired)
National Institute for Agronomic Research (INRA), Dept SAE2
Nantes, France
mzitt@numericable.fr

Michel Zitt received a PhD in Management (1979) and a Doctorate in Economics (1991). He mainly published on various aspects of bibliometrics and evaluation: citation measurement and normalization, science thematic mapping, hybrid word-citation techniques, and internationalization of science. Several works stress the limitations of indicators, dependence on data coverage, scale of observation and diversity of science. He was awarded the Price Medal in 2009.

Alain Lelu

ELLIADD (retired)
Université de Franche-Comté
Besançon, France
alelu@orange.fr

Alain Lelu received a PhD in Statistics and Computer science (1993). He published on Information Science, mainly on neural models applied to textual data, clustering, factor methods and intrinsic dimensions of data tables and graphs. He also wrote prospective papers on social impacts of telecommunications and data analysis.

Martine Cadot

Knowledge and Language Management
LORIA (Laboratory in Computer Science and its Applications)
Villers-lès-Nancy, France
martine.cadot@loria.fr

Martine Cadot received a MS in Economics (1995) and a PhD in Computer Science (2006). She taught Mathematics, Statistics and Computer Science. Her present research at LORIA focuses on building statistically robust models for real data, in collaboration with scientists of various domains. She participated in international knowledge discovery challenges (3rd place at KDD Cup 2002, 1st place at DEFT 2011).

Guillaume Cabanac

Computer Science Department
University of Toulouse
Toulouse, France
guillaume.cabanac@univ-tlse3.fr

Guillaume Cabanac is an Associate Professor at the University of Toulouse, France. He received a PhD (2008) and a habilitation degree (2016) in Computer Science. His research lies at the crossroads between Information Retrieval, Digital Libraries, and Scientometrics. He co-chairs the Bibliometric-enhanced Information Retrieval (BIR) workshop series and he serves on the editorial board of *JASIST* and *Scientometrics*.

3. Knowledge Integration: Its Meaning and Measurement

Ronald Rousseau, Lin Zhang, Xiaojun Hu

Interdisciplinary research depends on research traditions and fields originating from different research teams, different countries and regions. Its essence is knowledge integration. As a dynamic and interactive process it continuously pushes the structure of science to become a complex diverse system.

In this chapter, we provide a systematic review of interdisciplinary research. Starting from a definition of interdisciplinary research, its elements, and its role for scientific progress, we particularly focus on how to identify the activity of interdisciplinary research, how to measure it and point out the limitations of existing approaches. Stating that one can measure knowledge integration implies that this notion refers to a continuum, beginning from no integration (disciplinary research) to a large degree of integration (highly interdisciplinary).

Following Stirling, Rafols and Meyer we show that knowledge integration can be measured by two main factors: a diversity factor and a network coherence factor. The diversity factor itself consists of three aspects: variety (number of categories taken into account), evenness and similarity between categories. In accordance with the Jost–Leinster–Cobbold approach we prefer a so-called true diversity measure.

As an illustration, we provide a simple example of a study on interdisciplinarity in the field of synthetic biology, using the true diversity measure derived from the Rao–Stirling measure. Finally, we include some suggestions for future research.

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_3

3.1 Interdisciplinarity

Interdisciplinary research is generally seen as a source of creativity and innovativeness [3.1]. As a consequence, measures of interdisciplinarity serve as indicators for policymakers, research managers, evaluators and sociologists of science [3.2]. In order to simplify the notation we will often abbreviate the term *interdisciplinary research* by IDR and the term *interdisciplinarity* as well as *interdisciplinary* by ID. Although present research policy often implicitly assumes that IDR can readily be identified and tracked, this is far from true.

Providing policy makers with maps and measures that capture the intensity of ID or of knowledge integration is a scientific task of high practical importance, yet fraught with difficulties, see [3.2] for a review, or also [3.3].

As an example to show how official institutes talk about ID we mention the European Research Council's (ERC) grants aiming to support so-called frontier research, which describe frontier research as research of an interdisciplinary nature which crosses boundaries between different fields of research, addresses new and emerging fields and introduces unconventional, innovative approaches [3.4, p. 6].

It is important to be able to identify interdisciplinary activity, to know where it occurs and ensure it is properly assessed. However it should be recognized that these are actually three separate goals. It seems highly unlikely that each of these goals can be reached using the exact same methodology. Should indicators be adapted to the target group, i. e., policymakers, research managers, funding bodies, evaluators or sociologists of science, or is this not feasible in practice?

Adams et al. [3.5] mention that key policy topics—they looked at mental health and climate change—are supported by research drawn from a wide range of disciplines and conclude that this observation provides evidence to counter the view that only a few key disciplines are required to solve the grand challenges of our time. It seems common sense that major problems benefit from multiple perspectives. Yet, only when the degree of ID can be determined does it become possible to find out if they lead to the real *novel* results aimed for by governments and funders. How do we link *interdisciplinary research* and *novel research*? One surely needs precise indicators—or as precise as humanly possible—to support experts from the targeted scientific/technological fields.

Our work is based in large part on [3.2, 6–9]. We focus on quantitative approaches; qualitative approaches are mentioned only in passing.

3.2 Definitions

IDR can mean different things to different people. According to the National Academies of Sciences of the USA [3.10, p. 2] interdisciplinary research is:

> A mode of research by teams or individuals that integrates information, data, techniques, tools, perspectives, concepts and/or theories from two or more disciplines or bodies of specialized knowledge to advance fundamental understanding or to solve problems whose solutions are beyond the scope of a single discipline or area of research practice.

In this definition the key concept is knowledge integration. Knowledge integration may be realized by integrating knowledge originating from different disciplines (interdisciplinarity in the pure sense), but it may also happen that knowledge originating from different research traditions, different regions, from different ages or from different schools of thought is brought together. The more an article, or any other item under investigation, integrates different sources, the more it is interdisciplinary. Keeping these different integrative processes in mind we will, from now on, nevertheless focus on interdisciplinarity in the pure sense. Although some researchers make a distinction between the terms interdisciplinary, multidisciplinary, transdisciplinary and cross-disciplinary research [3.11] in empirical studies one finds a continuum which makes it difficult to distinguish among these modes [3.8]. Hence, we will use the term interdisciplinary as a general term, comprising the terms transdisciplinary and cross-disciplinary. We do see *multidisciplinary* as a separate term, referring in particular to journals which publish articles on different topics such as *Nature* and *Science*, which are typical multidisciplinary journals.

IDR is a phenomenon that emerges within and because of the dynamics of a larger knowledge system that includes external drivers, such as the need to solve complex problems and dependence on funding priorities [3.2, 12]. *Molas-Gallart* et al. [3.13] point out that there is a variety of IDR and that various processes through which research is defined, funded, conducted and applied have an effect on the extent and type of impact, social or otherwise, resulting from IDR. They state that IDR is often associated with problem-oriented research, as societal problems seldom conform to disciplinary boundaries, and with interactions that go beyond academia. Yet they admit that, on the one hand, there exist forms of IDR that do not address societal issues and do not interact with potential non-academic beneficiaries, as e. g., in biophysics. On the other hand they state that impact processes through which IDR can yield socio-economic benefits need not be unique to IDR.

It is obvious that one cannot use the term IDR if one does not first define the term discipline. Hence the term discipline logically precedes the term ID [3.14]. In this contribution we will use the term discipline as a synonym for research specialty, field or knowledge domain. Disciplinary structure can be captured using statistical tools, e. g., by applying a cluster algorithm, as in [3.15–17]; as a philosophical result (as the categories used in the universal decimal classification (UDC) or in the Dewey classification) or a practice-based categorization, supported by statistics, such as the subject categories of the *Journal Citation Reports* (JCR) [3.18, 19]. Once a categorization, with possibly arbitrary delimitations between disciplines, has been applied this might result in rigid boundaries which tend to hinder accurate descriptions of the dynamics of science. This inevitable tension between the use of rigid boundaries and the description of the dynamics of science is particularly acute in emergent fields [3.8]. This is one of the reasons why *Rafols* and *Meyer*, the authors of [3.8], try out a second, complementary network perspective that does not rely on pre-existing classifications. In this context using existing categories is referred to as a top-down approach, while starting from the data, leading to a network approach, is referred to as the bottom-up approach.

We further note that it was already included in the National Academies of Sciences' definition that knowledge integration can be realized by a single person. *Porter* et al. [3.20] define IDR as requiring an integration of concepts, techniques and/or data from different fields of established research, but do not presume the formation of teams. Knowledge integration is the focal point. Hence collaboration issues are of secondary importance in IDR studies. Yet, in the majority of practical cases IDR implies the collaboration of different persons, even of different teams. *Rafols* and *Meyer* [3.8] provide the example of a biophysics and a biochemistry lab collaborating in the field of bionanoscience. For this reason we will further assume that IDR implies a research team or the collaboration of two or more research teams. Moreover, we already mention that in our opinion collaboration is a strong form of knowledge integration and, most of the time, also of knowledge diffusion (in case one or more members of the team teach or explain to others what they know to advance the joint work) (Sect. 3.11).

IDR can be studied on the level of inputs, processes, outputs, and outcomes (Sect. 3.4). Yet in [3.21] the authors note that most observers do not specify which aspects or components they focus on. Besides on inputs, processes, outputs, and outcomes one may also focus on people, objectives, activities, and impacts.

Clearly, it is not possible to define a unique and absolute measure of IDR. For this reason several proxy indicators have been created. Each proxy indicator delivers different insights about the interdisciplinary nature of the research under study. *Adams* et al. [3.21] note that the same project may be indexed as interdisciplinary for one parameter (e. g., departmental affiliations or universities) and not for another (diversity of references). Hence, it is not surprising that different indicators may deliver inconsistent and even contradictory results. Considering diversity, of whatever type (see further), as a proxy of ID it is, however, always true that smaller diversity points to more specialized research, while larger diversity points to more integrative research.

For this reason *Adams* et al. [3.21] point out that it is essential to consider a framework for analysis, drawing on multiple indicators, rather than expecting any simplistic (their words) index to produce an informative outcome on its own. The main objective of [3.21] was to compare the consistency of indicators for ID and, if possible, identify a preferred methodology. Their study revealed that choice of data, methodology and indicators (their precise mathematical formulas) can produce seriously inconsistent results despite being applied to a common set of disciplines or countries. Their results highlight issues about the link (and how to clarify it) between any quantitative proxy indicator and the assumed target of the research about IDR. *Adams* et al. [3.21] even seem to imply that IDR almost always implies a policy target. Yet, we do not agree with this assumption which, in our opinion, seems to go too far.

We also note that when studying ID, cognitive differences among scientists working together may play an important practical role. For this reason we include in this chapter a section on measuring cognitive distance (Sect. 3.15).

Rafols and *Meyer* [3.8] use a concept of integration that reflects two aspects of knowledge systems:

(a) Diversity, itself consisting of three aspects, namely species diversity, evenness and disparity (explained and discussed further on)
(b) Coherence or the extent to which specific topics, concepts, tools and data used in a research process are related [3.2, 14].

For practical reasons disciplines are often operationalized as Web of Science (WoS) subject categories. Yet, it is well known that the delineation of WoS categories is far from ideal [3.22]. These categories are, moreover, overlapping, which is generally considered to be a disadvantage. This is one of the reasons why there is no all-preferred choice to represent disciplines, and other classifications besides the WoS subject categories are used. The classification scheme of the essential science indicators (ESI) contains 22 non-overlapping broad fields and was, for that reason, used in [3.23]. *Bromham* et al. [3.24] used field of research codes (FoR) as jointly developed by the Australian Bureau of Statistics and Statistics New Zealand in 2008. *Zhang* et al. [3.25] use Leuven-Budapest (ECOOM) classes [3.18] on two levels, namely that of its 16 major subject fields and that of its 68 subfields. One may also use categories which are not related to bibliometrics, such as medical subject headings (MeSH) categories [3.26, 27], the mathematics subject classification (MSC) [3.28, 29], the physics and astronomy classification scheme (PACS) [3.30, 31] or the international patent classification (IPC classes) [3.32, 33]. *Rafols* [3.9] points out that using practice-oriented categories might be especially helpful when analyzing the social impact of research, or the influence of one subfield on another [3.31]. He notes, however, that MeSH maps could not be matched with global maps of science, indicating that the underlying cognitive structure and metrics are different. With *Adams* et al. [3.34] we note that the degree of ID depends on how finely or coarsely categories are defined. The narrower the boundaries of disciplines are defined, the more likely it becomes that a piece of literature crosses borders between disciplines.

3.3 Drivers and Arguments in Favor of Interdisciplinary Research

Molas-Gallart et al. [3.13] distinguish four main aspects in which differences in IDR emerge:

(a) The primary objectives driving the research; here the authors make a distinction between challenge driven or societal challenges, and scientific/academic driven
(b) Cognitive distance between the bodies of knowledge bridged, or designed to bridge, through IDR
(c) The extent of integration among these bodies of knowledge (pre-existing or brought about by IDR)
(d) Practices by which researchers conduct IDR.

Interdisciplinary research teams consisting of several collaborating colleagues may lead to a combination of diverse perspectives, interpretations and models, and hence avoid ignorance (about certain aspects of the topic of research or about necessary technical skills) and lock-ins (doing what *everyone has always done*). In many cases specific scientific or technical skills must be combined with knowledge on human behaviour as, for instance, otherwise epidemics may never be conquered or new products may not reach those who might benefit. According to *Yegros-Yegros* et al. [3.35] the rationale for IDR is particularly strong and convincing in scientific programs addressing grand societal issues or challenges such as climate change, epidemics (AIDS, Ebola, Zika, SARS), and the preservation of ecological diversity. According to this line of thinking, project leaders with a mind-set belonging to *Stokes'* quadrant [3.36] in which Pasteur is placed, namely the quadrant combining fundamental understanding and practical use, are said to be the natural leaders of successful interdisciplinary work.

When it comes to practice an important question is about the kind of learning that is important in ID teams and the kind of information and knowledge exchange that occurs. Stated otherwise: *What kinds of exchanges form the basis of collaboration in ID groups?*. This has been studied in [3.37]. In this article Haythornthwaite asked members of teams what they thought the others learned from them. In this way she found that learning was not just about factual knowledge, but included learning about processes and methods, engaging jointly in research, learning about technology, generating new ideas, socialization into the profession, accessing a network of contacts, and administrative work. Obviously, more than just domain knowledge is shared in IDR. Distributions of these relations showed that there is more sharing of similar than of different kinds of knowledge, suggesting that knowledge may flow across disciplinary boundaries following lines of practice. Simply receiving information or observing materials without engaging with them is not sufficient for learning. It is the interaction with others that makes knowledge exchange particularly useful. In this, there may or may not be reciprocity.

It is also well-known that IDR entails costs [3.38]. In this contribution we will not go into details about these costs, but just mention that there are different kinds: those associated with team building and the coordination costs associated with it, and those associated with the lack of appreciation of IDR, in the sense that academic structures are still largely based on monodisciplinary structures. For this reason scientists performing IDR often encounter problems in getting tenure or other academic rewards. Besides lack of institutional appreciation we also mention a general bias in evaluation as evaluation standards are often disciplinary based [3.39, 40]. *Bromham* et al. [3.24] found that whatever the general perception, IDR was less funded than non-ID research. Recently *Leahey* et al. [3.41] found another *penalty* for doing IDR, namely lower productivity, yet combined with increased citations (more detailed information on a possible increase in citations is provided in Sect. 3.14). They conclude that IDR is a high-risk, high-reward endeavor.

3.4 Different Aspects of Interdisciplinary Work

The aspects reviewed in this section refer mainly to inputs, processes, outputs and outcomes of interdisciplinary work.

3.4.1 Inputs

Inputs for IDR projects, whether or not they start as such or become ID during their development, include antecedents (what exists before the project starts), such as teams, team building capacities of the project leader and other personal factors of researchers within a research group. The physical environment, as well as bureaucratic and structural issues such as existing policy documents can also be considered as inputs to IDR [3.42]. In the initial stage of an IDR project team building and team expansion are important parts which can be described in a qualitative way.

When IDR is about providing practical solutions for concrete problems of a local community, i. e., a local community is the final targeted beneficiary, then representatives of this community should be included from the start [3.43], and as such should be considered as part of the input.

When planning/starting an ID project several aspects must be taken into account, among which required skills and needed equipment are two essential ones. Hiring new researchers and technicians with specific skills may be necessary, as is reaching out to another team to join forces (for skills and/or equipment). Yet, many projects not initially conceived as such evolve to become IDR when it is realized that more insight into the research question is required.

3.4.2 The Process Itself

Knowledge integration is a process. Like any other process it has inputs, outputs and outcomes. Yet, here we consider the integration activity itself. This activity can be described when it happens or can be reconstructed afterward. Both forms are qualitative and often participative. *Sanz-Menéndez* et al. [3.44] used a survey technique to reach what they refer to as a comprehensive insight in the degree of interdisciplinarity in three research areas:

- Pharmacology and pharmacy
- Cardiovascular systems
- Materials science.

Their survey included questions related to socioprofessional features of individual researchers, the cognitive context of references, team features and external collaboration. They concluded that during IDR the aspects of specialization, fragmentation and hybridization all come together.

Smith [3.45] wrote that citations can be seen as historic remains left by interactions of different ideas related to the object of study. For this reason we claim that article references are remnant signs of the process of knowledge integration as it has happened during investigations. Hence, studying ID can be done through a study of article references.

3.4.3 Outputs

There is no special output format which is typical for IDR; articles in journals and edited books, monographs, and patents can all be used. Yet, an edited book or a special issue of a journal could be mentioned here, when contributions are written and edited as the result of an ID project. This observation holds in particular if editors have, more than usual, taken care that the contents of chapters are highly integrated and each chapter presents unique aspects of integrative work.

Several outcomes are highly desired and typical, be it not unique, for IDR. These are discussed next.

3.4.4 Outcomes

The outcomes we describe can be reached through single-authored work, a mono-disciplinary team in one department or through multidisciplinary work. Yet, we mention them here as they are often described (when evaluating IDR) or are aimed for in IDR. These outcomes may have a direct or an indirect impact on academia and/or society:

- Dissemination of project results reaching the intended audience(s).
- Advising policy makers.
- Citations in the general literature, especially when these occur in a variety of disciplines (knowledge diffusion, Sect. 3.11). New knowledge injected into other fields is one possible outcome of IDR. This aspect can be studied using diffusion indicators. In medical studies the desired outcome is often an increase in public health. An intermediary outcome could be the influence on medical advisory committees and in medical guidelines [3.46].
- Setting up websites as information tools.
- Dedicated meetings with stakeholders.
- The ability to work across the academic/non-academic boundary can be seen as a capacity-building impact.
- Successful translational research in which findings from basic science really enhance human health and well-being through life saving point-of-care patient applications; or the reverse direction in which clinical findings lead to new basic insights.
- Innovative ideas resulting from ID reflections, see Sect. 3.16.
- Being able and prepared to respond to support excellence and respond to new opportunities. According to [3.47] this ability comes from:
  - Diversity in research fields: A broad range of disciplines supports exceptional levels of research excellence.
  - Diversity in support which gives flexibility of research support to allow a mix of long and short terms responses and includes strategic and responsive awards. Diverse funding mechanisms are required to enable curiosity-driven research and evolving targeted programs of high policy priority or scientific need.
  - Diversity of research organizations, where mission-led units complement large and small universities with regional as well as international engagement. *Robert May* [3.48] showed that research economies with a strong university research base performed consistently better than those committed to narrow, mission-led research institutes.
- Supporting emerging new fields (such as synthetic biology, translational medicine).

3.5 Quantitative Measures: Introduction

In this section we begin our discussion on how to measure the degree of interdisciplinarity. Stating that one can measure the degree of ID implies that this notion refers to a continuum, beginning from no disciplinary research to a high degree of ID. Before going into details, we mention here that in this section we will not cover the evaluation of IDR. As we consider this a totally different issue, we will discuss the evaluation of IDR in another section (Sect. 3.13).

Which databases are best used in IDR? The WoS and its sub-databases Science Citation Index (SCI), Social Sciences Citation Index (SSCI) and Arts & Humanities Citation Index (A&HCI) are often employed. Yet, this use can be considered as an historical path dependency, in the sense that colleagues tend to collect data from the same databases as their predecessors. If IDR involves a large-scale study then databases that provide support for bulk download or include metadata that support IDR analysis are preferred [3.2]. We note, though, that not all studies of IDR are large-scale studies; for instance *Rafols* and *Meyer* [3.8] performed a detailed study of some case studies in one particular field (bionanoscience).

Wagner et al. [3.2] make another distinction, namely between approaches that account for the larger system and those that do not. Those that do, do so by viewing the science system as a whole. They use statistical relationships among key aspects of relevant or desired data to measure desired characteristics of aggregation of authors, articles or journals.

3.5.1 Top-Down (Classification-Based) and Bottom-Up Approaches

Most current bibliometric-based measures of IDR input and processes rely on the journal categories established by the Institute for Scientific Information (ISI). However, the use of journal categories for measuring IDR has some obvious limitations because of the dependence upon a predefined taxonomy or category structure. Classification-based measures can be useful as a first start, especially when used to compare large areas of science with large amounts of data (an application of big data analytics). Nevertheless, issues around the use of underlying taxonomies or classification schemes as the basis for IDR measures makes this approach problematic. With no consensus on the best categorization and considerable evidence that various measures of IDR yield quite different results depending on the classification system chosen for analysis, it is apparent that problems remain for any IDR measure based on a predefined classification scheme [3.2]. To these objections we like to add that ID should be studied on the article level: journals are, in general, not an appropriate unit.

Bottom-up approaches based on network clusters formed by articles, using co-citation, co-word or bibliographic coupling analysis, can capture knowledge integration in the making evidenced by people working on new problems [3.2]. We note that this statement implies a time aspect, see further Sect. 3.17. We also note that the statement about "knowledge integration in the making" is certainly not true for co-citation analysis, as it takes too much time to gather enough citations to lead to a meaningful timely analysis.

Cassi et al. [3.49] calculate the degree of interdisciplinarity of an institution. As such they need to combine ID values of individual articles published by members of the institute with the ID value of the institute as a whole. Is the ID of an institute the average of the ID of its articles, or should ID be measured directly from the set of all articles published by the institute? To answer this question they applied a decomposition of the Rao–Stirling measure (this measure is explained further on). This means that the diversity of a research institute is the sum of the diversities within each article it published, plus the diversity between articles. This decomposition is performed in terms of inertia of sets of points. These authors further provide an interpretation of their results by comparing with a benchmark, namely the value of their index calculated for the world production in the same field and the same period of time. They then perform a statistical test to decide whether a difference is significant or not.

Wagner et al. [3.2] make a distinction between two approaches: The structural approach and the spatial approach. In the next section we first consider the structural approach, leaving the spatial approach to Sect. 3.7.

3.6 Structural Approach

According to [3.2] the structural approach is better at capturing emerging developments that do not fit into existing categories, while the classification-based approach might be useful at large-scale explorations such as the disciplinary breadth of universities. In a time where many universities strive to be *of world class*, studying the disciplinary breadth, combined with ID aspects, is an interesting idea that should be elaborated further.

As *Zitt* [3.50] points out, the literature does not converge on a universally accepted methodology or basis from which to uncover the structure of science, let alone how that structure may reveal IDR. Structural diversity is the diversity of disciplines, institutions and support mechanisms. It is a property of a research base that has the capacity to address the challenges of tomorrow, and to do this in a flexible and responsive way [3.47]. It is distinct from social diversity, i. e., diversity in gender, nationality, and ethnicity.

Most ID studies apply citation analysis. Yet, methods of text analysis and hybrid methods are other options [3.51]. *Rafols* and *Meyer* [3.8] note that the percentage of references outside the discipline of the citing paper is a simple and often used indicator of IDR. This method was used e. g., in [3.52, 53]. This is also the method suggested by *Garfield* et al. [3.54]. When studying the ID of a set of documents, such as the publication output of an institute or research group, one may focus on the publications themselves and their distribution over categories, or one may examine article references or one can take into account the categories of articles citing the articles under study [3.49]. We think, however, that considering citing articles is an aspect of diffusion and not of knowledge integration.

Citation analysis privileges publications as the major outcome of IDR. This is the most important limitation of this approach. This leads to the question [3.2]: in relying on quantitative, publication based measures, what information about IDR impact is lost?

Any measure that includes citations/references has to deal with the problem of *why does one cite?*, including preferences for certain journals and/or languages. Citation analysis is, moreover, complicated by the fact that disciplines are not sharply defined and are vastly different in size. For this reason relative measures, applying normalization, are needed. But when applying normalization the properties of various normalization techniques must be understood, explained and taken into account, which is not a sinecure. Studying disjoint units makes measures clear and precise. Although this is a clear advantage it is a fact that IDR is about overlap between fields; hence overlapping study units, and not disjoint ones, seem unavoidable when studying IDR.

Zitt [3.50] offers a cross-scale perspective to address the aggregation problem in order to make indicators more comparable across fields, while [3.55] presents a macro-approach to IDR at the article level to test the impact of the level of aggregation (measured by the number of citations to a unit of output). The authors of [3.25], too, include a comparison with respect to levels of aggregation.

Analysis of IDR at the journal (citation) level is performed by placing journals into disciplinary categories and then viewing the extent to which they have relationships with journals outside that disciplinary category. In this context [3.56] and [3.57] make a clear distinction between the cited and the citing side: measuring IDR at the journal level as either an input (from the cited side, measuring its contribution to or impact on other journals/disciplines) or an output (from the citing side, measuring its uptake or use of information from other journals/disciples). The set of relationships between journals can be represented in tables or in an abstract space based on vector analysis, showing similarities or linkages and differences or variations. This process brings our discussion to the point of considering a metaphorical knowledge space in which to conduct IDR analysis, see Sect. 3.7.

If one knew the field/discipline of each author, then co-authorship could be used to describe IDR. In this approach, as mentioned in [3.58], the problem of interdisciplinary recognition shifts from the semantic analysis of an article or the scientific classification of cited (or citing, in the case of diffusion) papers to the identification of its authors' specialization. A practical way of determining an author's field is through their departmental or institutional affiliation. Yet, departments or institutes are organizational units, not disciplinary ones. Another obvious problem is the case of interdisciplinary institutes not focusing on one particular field. *Abramo* et al. [3.58] had the exceptional advantage that in Italy each academic scientist classifies him/herself in one and only one scientific field, 370 in all. These fields are grouped into 14 disciplinary areas. Also [3.12] used the distribution of researchers over a set of subdisciplines. Clearly, as long as scientists regularly update their field (if necessary) this looks like an ideal situation for studying IDR.

3.7 IDR in the Research Landscape

A second approach to using bibliometrics is a methodology that describes a landscape or space within which science operates [3.59]. *Wagner* et al. [3.2] note that spatial distance as a measure to analyze and visualize IDR was already suggested in [3.54]. The practical usefulness of this approach has been enhanced by recent developments in computing and algorithms that can be used to standardize the analysis and to bring into view the underlying dynamics of the relationships among and across disciplines. Programs such as Pajek [3.60] and VOSviewer [3.61, 62] have important applications in ID studies. Visualization in overlay maps is a way of providing a description of diversity and coherence without the need to collapse the data into a single figure [3.9].

A classical method is to use factor analysis in which each factor can be labeled as a category or discipline. Then the elements of the matrix that load multiple factors above some preset threshold are considered the most interdisciplinary [3.63, 64]. Factor analysis is an example of an approach that defines an emergent structure from data based on a set of distances, rather than a pre-imposed structure.

3.8 Concrete Measurements

In this section we come to mathematical formulas to measure ID. We will focus on the measurement of diversity, through its three components:

1. Variety
2. Balance
3. Disparity.

Wagner et al. [3.2] discuss concrete measurements but do not provide precise mathematical formulas, hence cannot explain why some indicators are, for logical reasons, better than others. With *Jost* [3.65] we stress that one must not confound a concept (here interdisciplinarity and diversity) with an index that measures it.

Morillo et al. [3.66, 67] attempt to establish a tentative typology of disciplines according to their degree of interdisciplinarity measured through multi-assignation of journals to WoS subject categories by which the assignment of a journal to more than one category indicates the existence of cognitive links between disciplines, which can result in interdisciplinary research. Note that the word *can* is of importance here, since such links may also be the result of the multidisciplinarity of these journals.

Several papers suggest combining similarity and distance measures to build an indicator. In a seminal paper *Porter* et al. [3.20] developed measures of integration and specialization of research outputs to understand where IDR is situated within the system of science. Their integration indicator measures cognitive distance among the subject categories in which a body of research is published. Somewhat similarly, [3.68] suggests a diversity index as a way to incorporate multiple measures of IDR. Their index measures how particular research articles integrate research fields, based on the assignation of the journals they cite to WoS subject categories. The line of thought that began in [3.20] was taken up again in [3.8], which we will discuss in more detail. Measures using similarity and variation are also used to compute network properties such as betweenness centrality, an application of which was proposed in [3.56].

After having reviewed a wide range of documents on diversity measurement *Stirling* [3.69] concluded that diversity consists of three basic concepts: variety, balance and disparity, each of them being a necessary but insufficient property of diversity as a whole. Species in ecology, WoS categories in informetrics or forms of energy (another example considered by *Stirling* [3.69]) are not independent entities, but they are shaped by patterns of common development or ancestry leading to proximities (or the opposite: disparities) between units. The key of an acceptable integration score is that it captures not only the number of disciplines cited by a paper and their degree of concentration but also provides a measure of how disparate these disciplines are. In order to do so, it relies on a specific metric of distances or similarities between pairs of disciplines. *Stirling* [3.69] concludes that the Rao–Stirling measure, which can be interpreted as a distance-weighted Simpson diversity, is such an acceptable measure. It is defined as

$$D = \sum_{\substack{i,j \\ i \neq j}} (d_{ij})^{\alpha} (p_i p_j)^{\beta} \; . \tag{3.1}$$

Here d_{ij} denotes the dissimilarity (disparity) between category i and category j, and p_i and p_j denote the proportions of the total number of items under study in category i and category j, respectively; finally α and β are parameters that adjust the importance given to distances among categories (α) and proportions (β).

Next we provide a short overview of elements taken into account in IDR studies.

Here, the systems or units of analysis can be university departments, an emergent topic represented by a set of articles (hence from different departments all over the world), researchers, research teams or a funded project.

Elements can include articles, monographs, dissertations, projects or funding proposals.

Elements must be classified into categories. The most straightforward way of assigning bibliographic elements, such as articles or references to categories, is to rely on the categories provided by databases providers. The most widely used classification is Clarivate Analytics' (formerly ISI and later Thomson Reuters) Web of Science categories which is journal-based (not article-based) and very problematic, as articles published in the same journal do not necessarily share a similar topic or research perspective [3.9]. An alternative is to use a bottom-up approach in which investigators themselves carry out a clustering of articles based on citation patterns. This may lead to disjoint categories corresponding with research practices. A probably better approach is to cluster articles based on keywords. In this way it becomes possible to obtain clusters which include emergent topics and fields. Yet, it is obvious that such clusters change in a dynamic way (they are born, die, divide) [3.70] and hence such investigations may be difficult to replicate, especially if underlying data may not be made public.

Providing an estimate of the cognitive distance, the d_{ij}-values in the disparity component, between categories is challenging. Global maps of science as developed in [3.19, 22, 26, 57, 71–76] may be useful here. Yet, as suggested in [3.77] we think that a direct measurement is better than an indirect measurement via these maps of science. In our opinion, such maps should be used only as visual illustrations.

A table showing examples of different choices of systems, elements, categories and metrics to measure diversity is provided in [3.9].

3.9 Entropy is not the Same as Diversity or Interdisciplinarity

As the notion of entropy is often used in the context of IDR [3.12] and of measuring diversity we take some time to explain this concept and its difference with the notion of diversity. Entropy is a mathematical concept drawn from statistical thermodynamics (Clausius being the originator), statistical mechanics (Boltzmann) and information theory (Shannon). It is a measure of disorder or uncertainty in a system, in this case the system of science. The entropy measure is a particular case of a measure used in practice to describe diversity [3.8, 69, 78].

Jost [3.65], however, makes a strong point showing that entropy and diversity are different notions which should not be confounded. We repeat his arguments but instead of a biological context we embed them in an ID context. Concretely, we want to measure ID by considering references and the spread of references over disciplines. Consider an article with A references originating from E equally cited disciplines. Hence each discipline is used A/E times. It is then reasonable to say that this article has an interdisciplinarity of E disciplines. If now these A references originate from $2E$ equally cited disciplines, then this article has an interdisciplinarity of $2E$. If all cited disciplines are equally common then ID is proportional to the number of cited disciplines. It is, moreover, natural to set the proportionality constant equal to one. Considering now the entropy measure

$$H = -\sum_{j=1}^{N} p_j \log_b(p_j) \;, \tag{3.2}$$

where p_j denotes the fraction of items assigned to category or cell j. If $N = E = 4$ and taking $b = 2$ yields

$$-\sum_{j=1}^{4} \frac{1}{4} \log_2\left(\frac{1}{4}\right) = 2 \;, \tag{3.3}$$

for the first case and

$$-\sum_{j=1}^{8} \frac{1}{8} \log_2\left(\frac{1}{8}\right) = 3 \;, \tag{3.4}$$

in the second case. Hence, one would conclude that ID has not doubled. The reason is that entropy is related to the uncertainty of finding a discipline in the sample, not to the number of disciplines. Similar counterexamples can be obtained for other indicators such as the Gini index. Each ID index creates equivalence classes in the set of all study objects, here reference lists. All reference lists leading to the same value of the indicator are considered to be equivalent. In each equivalence class there is exactly one case where all disciplines are equally common (admitting any real number as the number of disciplines). Hence, finding the ID of a reference list, starting from any ID measure reduces to the problem of finding an equivalent reference list composed of equally common disciplines. Recall that if there are E equally common disciplines, then each occurs with a relative frequency of $1/E$.

Given an article with differently cited disciplines with proportions $(p_j)_j$, its entropy value H_0 is

$$-\sum_{j=1}^{S} p_j \ln(p_j) \;. \tag{3.5}$$

The problem is now to find the number E (the number of equally common disciplines) for which the entropy value, denoted as $H(E)$ is equal to H_0. Now

$$H(E) = -\sum_{j=1}^{E} \frac{1}{E} \ln\left(\frac{1}{E}\right) = \ln(E) \;. \tag{3.6}$$

Hence

$$\ln(E) = H_0 \text{ or } E = \mathrm{e}^{-\sum_{j=1}^{S} p_j \ln(p_j)} \;. \tag{3.7}$$

This E is the so-called true interdisciplinarity for the article with different cited disciplines with proportions $(p_j)_j$. It is now clear that there is a considerable advantage in using ID measures which coincide with their true interdisciplinarity. Indeed, only then one may rightly say that a value $2y$ refers to a double ID as the value y.

We conclude that there is nothing wrong with the idea of using entropy for diversity measurement, but instead of the classical Shannon entropy measure, one should use its true ID analogue.

3.10 The Rafols–Meyer Framework

The main purpose of the article written by *Rafols* and *Meyer* [3.8] was to investigate if, indeed, more ID leads to better research. In order to answer this question two other questions must be answered. The first is to define scientific performance, i. e., excellence in scientific research, while the second is to define interdisciplinarity, more precisely the intensity of IDR, as this is not a yes-no question. *Rafols* and *Meyer* [3.8] do not study the first question but propose a methodology to answer the second one. After mentioning that the key concept is knowledge integration, they come to the conclusion that two aspects must be investigated: diversity and coherence. In this context the term diversity relates to high cognitive heterogeneity, while coherence relates to a process in which previously different and disconnected bodies of research become related. Coherence emphasizes how different bodies of research are consistently articulated and form a meaningful constellation [3.8]. Diversity in this context is disciplinary diversity while coherence is a network property. Following [3.69] the notion of diversity consists of three aspects:

- Variety: The number of distinctive categories, here disciplines
- Balance or evenness in the distribution of categories
- Disparity (or its opposite: Similarity): The degree to which categories are different/similar.

All these notions must be operationalized and a concrete measure must be chosen. In a first step Rafols and Meyer choose the unit of investigation, namely separate articles. This means that they want to investigate to what extent (recall that they study the intensity of IDR) an article can be said to be ID. The next choice is to consider the references of this article. This means that, as stated above, they consider these references as the remnants of the knowledge integration process that has taken place during the research reported in the article.

Diversity and coherence are related to the stage of a research field. They can be combined, leading to four possible combinations [3.8], see Fig. 3.1:

- Low diversity and high coherence: All the references are from the same discipline and are highly related. This is the case of specialized disciplinary research.
- Low diversity and low coherence: The investigated article studies distant research specialties within the same discipline.
- High diversity and high coherence: References come from many disciplines, but are similar. This suggests that the article is the result of a completed interdisciplinary effort. Although it is highly interdisciplinary there is no new knowledge integration.
- High diversity and low coherence: References originating from many disciplines that were hitherto unrelated. This is from our point of view the most interesting case, as it indicates potential new knowledge integration.

Knowledge integration increases if diversity and coherence increase, or at least one of the two increases and the other stays invariant. Now further steps in the operationalization process must be taken. As disciplines, Rafols and Meyer use WoS subject categories (SCs) of

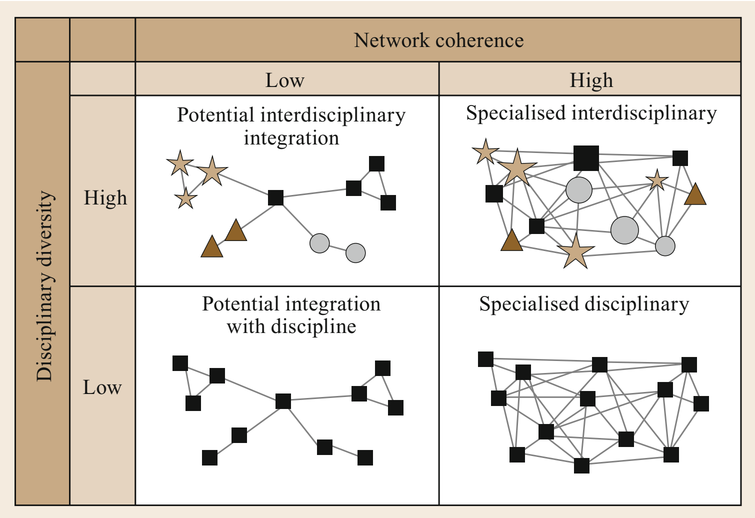


Fig. 3.1 Symbolic relation between diversity and coherence (after [3.8])

the references of references, but they note that if the initial set is large enough diversity can be computed directly from the SCs of the references. Coherence is operationalized through the bibliographic coupling network of the references. This allows the network relationship to reveal the structural consistency, or lack thereof, of the reference network. These choices imply a top-down approach for diversity as the items are classified according to a given categorization, and a bottom-up approach for coherence as the network is constructed from the data.

The final step is the choice of indicators. Rafols and Meyer calculate variety, the Shannon entropy, the Simpson and the Rao–Stirling measure as diversity measures. They decide on the Rao–Stirling measure as their final choice, as this measure takes disparity into account. We recall that the general Rao–Stirling measure is defined as shown in (3.1). In case one lacks empirical reasons to adjust α and β, they are often taken equal to one, which is also done by *Rafols* and *Meyer* [3.8]. Moreover, if two categories are identical or near-identical (distance zero or almost zero) then the sum of their separate contributions to the final diversity value is (almost) the same as the contribution of the single category consisting of the union of the two. This is a good property as it protects, at least partially, against misclassification.

Coherence is calculated as mean linkage strength, i. e., the mean degree centrality normalized by network size, and mean path length, i. e., the mean of the closeness centrality after dichotomizing similarities. These two measures turn out to be highly correlated. The authors decide on mean linkage strength as their final choice. This leads to two indicator values for each article under study. Values for each article are shown on a two-dimensional map, illustrating the uniqueness of each result. Yet, values of diversity as measured through the Rao–Stirling measure are rather similar in their case studies.

The framework used by *Rafols* and *Meyer* [3.8] separately analyzes the two key concepts necessary for the definition of knowledge integration. Diversity, on the one hand, describes the existing differences in the bodies of knowledge that are integrated, and coherence, on the other hand, describes the intensity of the relations between these bodies of knowledge [3.9]. *Rafols* and *Meyer* [3.8] suggest that the combination of these two approaches may be useful for comparative studies of emergent scientific and technological fields where new and controversial categorizations are accompanied by equally contested claims of novelty and interdisciplinarity [3.79]. In [3.9] the author proposes to subdivide the notion of coherence into three aspects: density, intensity and disparity. We are, however, not convinced about the added value or even feasibility of this approach and will not include this aspect in our investigations. We leave an in-depth investigation of this problem for others.

The Rafols–Meyer approach can be extended to patents, using global maps of technology and related measures of diversity. Moreover, the framework can also be applied to the integration of organizations, cities or countries [3.80]. Besides applying this framework to cognitive distance, one may try to apply it to the other four proximities discussed by *Boschma* [3.6] (see Sect. 3.15 for details). Extending this framework to other dimensions, it would be possible to investigate how knowledge integration is mediated by geographical, organizational, institutional and social networks [3.81].

3.11 Knowledge Diffusion as the Mirror Image of Knowledge Integration

We already mentioned that new knowledge injected into other fields is one of the possible outcomes of IDR. In this section we delve somewhat deeper into this diffusion aspect.

Since collaboration implies a direct form of knowledge sharing it can be considered a strong method of knowledge diffusion [3.82]. Knowledge sharing, and hence diffusion, through collaboration takes less time than knowledge diffusion through reading and citing. During collaboration, knowledge diffusion is nearly instantaneous, since during the joint work authors share knowledge and expertise. This happens even before a paper has been written. The knowledge shared between authors diffuses over the research groups, departments, institutes and countries to which the collaborating authors belong. This type of knowledge diffusion can be placed in the ID process, and consequently influences outputs and outcomes.

3.11.1 Knowledge Diffusion as a Property of One Article

Contrary to the case of knowledge integration, knowledge diffusion with respect to one article is largely determined by outsiders [3.14]. In [3.14] the diffusion process and related measurement has been described in some detail. The authors introduce the concept of an intermediary (IM) set. This is a set placed between the objects of inquiry, e. g., articles and objects characterizing the type of diffusion under study. Concretely, this may be disciplines when studying diffusion over disciplines, characterized by WoS subject categories; or journals, when studying diffusion over journals, etc. For the knowledge integration case studied in [3.8], the object of inquiry was an article; their aim was to study knowledge integration (interdisciplinarity) characterized by WoS subject categories and the IM was the set of references-of-references. If one wants to study how the knowledge contained in one article is diffused over journals, then the IM consists of citing authors acting through citing articles, which are each mapped to the journal in which the citing article has been published.

3.11.2 Interdisciplinary Knowledge Diffusion as a Property of a Set of Related Articles

Everything we have discussed so far about single articles also applies—with some adaptations—to groups of articles. However, changing the focus from one article to a group of related articles leads to an interesting new aspect.

The act of publishing subsequent research, e. g., written by the same research group, leads to a form of diffusion by publication. This has been pointed out in [3.23]. In this context diffusion as described above may be described as citation diffusion, i. e., diffusion by being cited. If a group publishes in different journals, or even different (sub)fields this may or may not be a token of knowledge integration. Whether knowledge integration has taken place must be determined by the methods described above: there is integration if the publication pattern is diverse and coherent, but not if the diversity is not articulated as a coherent body of research. This latter case is what occurs in multidisciplinary contexts, for example in centers that put diverse scientific groups under the same roof without achieving links between them—or in multidisciplinary journals such as *Nature* and *Science*.

As diffusion through citations is determined by citing articles it is a measure of reach [3.14]. If one is interested in a measure determined completely by outsiders, self-citations must be removed. Diffusion is largely determined by factors outside the original set of articles (or book, etc.). Superficially, the notion of knowledge diffusion has little to do with the article itself (or articles themselves), as it is more concerned with the question of how this article is received by the scientific community. Yet, the reason or reasons why an article is highly diffused has to do with intrinsic properties (intellectual and other) of the article itself. Publication diffusion is determined by the group of authors under study but is different from knowledge diffusion.

Diversity can be understood in the same way for knowledge integration as for knowledge diffusion. Coherence, however, has another interpretation when considered in the context of integration or in the context of diffusion. More coherence means more integration, but it may mean less diffusion, in the sense that the knowledge is spread over more but related topics. If diversity increases then clearly knowledge diffusion took place. For the relation between coherence and knowledge diffusion the answer is less clear.

3.12 Other Network Measures

Leydesdorff [3.56] suggested that betweenness centrality can be used as a measure of interdisciplinarity at the journal level. Using network measures in studies of IDR often means focusing on the intermediation role of certain actors, say some specific scientific contributions. *Rafols* et al. [3.83] developed intermediation as a framework, complementary to the diversity-coherence framework.

Using factor analysis, *Leydesdorff* and *Rafols* [3.19] analyzed the full set of ISI subject categories to find that it can be decomposed into 14 factors or macro-disciplines (as a name or an interpretation). Each of these macro-disciplines can be clustered, with the clusters providing spaces of high population (density) amidst open spaces. In their vision, links across these spaces show the interdisciplinary character of the populated spaces. This system-wide mapping of population spaces and their links provides a basis for determining the extent to which different fields are more or less interdisciplinary in character. Yet, the problem with this approach is that it depends on the classification used (WoS categories), though it has been shown [3.73] that most global mapping studies agree on the core structure of science. This suggests that although at the fine level the specific science map may be dependent on the classification, at a coarser level these differences are not relevant.

3.13 Evaluating Interdisciplinary Work

Klein [3.7] emphasizes that appropriate evaluation of IDR evolves through a dialogue between conventional and a new, expanding group of indicators for quality. According to Klein, traditional methodology and statistics have a role to play, but they are not sufficient to fully reveal the underlying dynamics. Describing time-dependence through dynamical aspects is, in her view, essential for an appropriate evaluation of IDR.

In [3.7] *Klein* considers seven generic principles, which we will review next, providing a coherent framework for thinking about the evaluation of interdisciplinary work:

1. Variability of goals. It must be accepted that IDR is performed with different goals in mind. Some teams focus on workability and social impact, while others seek simplicity or predictive power. Still others want to reach new levels of comprehensiveness or empirical grounding.
2. Variability of criteria and indicators. One set of criteria consists of conventional metrics, such as number of publications, citations or patents. Yet, another set emphasizes feedback to multiple fields, expanded expertise and tool sets, and the ability of researchers to work in more than one discipline, and as such being able to co-mentor students in other than their original field of expertise.
3. Leveraging of integration. In IDR evaluation one not only calls attention to outcomes but also to the quality of the process. Klein mentions that the Harvard project [3.84] highlighted the epistemic criterion of balance in weaving perspectives into a coherent whole. Recall that coherence is one of the two main aspects studied in [3.8].
4. Interaction of social and cognitive factors in collaboration. Clearly IDR is a special type of social process, one leading to knowledge production and integration. During collaborations, differences in opinion or in approaches to attack problems must be negotiated to avoid misunderstandings and to strengthen conditions leading to a consensual mode of work. Hence, communication and negotiation are important points in the formative evaluation of research projects. In this context we mention that having an open communication and management structure is one of the points in the evaluation procedure for gauging the performance of key labs in China [3.85]. ID teams form a network, but they still stay connected to the networks they belonged to before the team was formed. While the out-team connections may help in bringing new information into the network, they may also compete for individuals' time and attention. This suggests that out-team as well as in-team connections should be examined when assessing team dynamics and performance [3.37].
5. Management, leadership and coaching. Competence, especially of IDR leaders, is at least in part defined in terms of how well the management of projects and programs implements consensus building and integration. Leaders must perform cognitive, structural and process-oriented tasks and act as coaches, a task related to the topic of team building.
6. Iteration in a comprehensive and transparent system. Feedback must lead to new and improved methods and models. Transparency requires that evaluators as well as participants in IDR are from the outset informed about evaluation criteria.
7. Effectiveness and impact. Unintended (positive) consequences and unforeseen long-term impacts cannot be captured by a priori criteria. *Boix-Mansilla* et al. [3.84] mention that interdisciplinary impacts are often diffused, delayed in time and dispersed across diverse areas. This implies that a full evaluation of an IDR project may need a broad perspective of more than one decade.

Point 7 brings us to the question of whether IDR always leads to more citation impact.

3.14 Does Interdisciplinary Research Have More Impact?

First we point out that having an impact is different from receiving many citations. *Meagher* and *Martin* [3.86] point out that the impact of a field like mathematics—outside its own field—is largely realized through active collaboration with colleagues from other departments, i. e., through interdisciplinary research. That said, we next consider the topic of citation impact.

Molas-Gallart et al. [3.13] point out that increasing impact and increasing interdisciplinarity are not systematically positively correlated, a point already made in [3.53]. Also [3.35] warns against the idea that higher citations may reflect all benefits of IDR. They mention that, for instance, the aspect of opening up perspectives is not directly measured by high numbers of citations. Also the aspect of problem solving as a result of IDR may be underestimated, in particular when these solutions are associated with practical applications [3.87].

The question of whether articles that show a higher degree of knowledge integration are more cited is a complex question. In the context of interdisciplinarity, the answer seems to be heavily dependent on the definitions and measures used to gauge ID, and the field normalization used to count citations. Clearly the answer is certainly not a straightforward yes [3.35, 53, 88, 89].

Standard indicators may not be adequate to reflect the impact of IDR. For instance on the institutional level, journal rankings, being mostly discipline-oriented, can disadvantage diverse, interdisciplinary research in research evaluations [3.83]. On the country level, we note that high average citation impact hides a mix of peaks and troughs. More consistent performance, i. e., greater evenness in relative citation impact, avoids the risk of missing key areas [3.47].

Yegros-Yegros et al. [3.35] studied the difference on impact of proximal and distal interdisciplinarity. Citation impact is operationalized in terms of number of citations after field-normalization. In their work interdisciplinarity is measured by variety, balance and disparity of references. They use a tobit regression model to examine effects on citation impact, controlling for other variables such as number of authors and organizations. They observe an inverted U-shape relationship between degree of interdisciplinarity and citation impact. Variety has a positive effect on impact, while balance and disparity have a negative effect. These results indicate that the most-cited publications are those with a clear disciplinary focus, but that nevertheless give small proportions of references (the low balance) to many proximal (low disparity) disciplinary categories.

These findings can be interpreted in two different ways. First, while combining multiple fields has a positive effect on knowledge creation, successful research is better achieved through research efforts that draw on a relatively proximal range of fields, as distal interdisciplinary research might be too risky and more likely to fail. Yet, a second possible conclusion is that scientific audiences are reluctant to cite unorthodox papers that mix highly disparate bodies of knowledge. This would put publications that are too ground-breaking or challenging at a disadvantage.

Also *Wang* et al. [3.89] studied the influence of variety, balance and disparity separately. They found that long-term (13 year) citations increase at an increasing rate with variety, decrease with balance, and increase at a decreasing rate with disparity. Furthermore, although variety and disparity have positive effects on long-term citations, they have negative effects on short-term (3 year) citations, and although balance has a negative effect on long-term citations, its negative effect is insignificant in the short run.

3.15 Measuring Cognitive Distance

The capacity of sharing knowledge, possibly leading to new discoveries, depends on the cognitive distance between members of a team. *Broström* and *Mc Kelvey* [3.90] describe cognitive distance as an inverse characterization of the degree of overlap between two nodes in a semantic network in terms of knowledge bases, values, norms and the heuristics of attribution and decision making. *Boschma* [3.6] sees cognitive proximity, the opposite of cognitive distance, as one of the factors that facilitate effective collaboration between different actors in translational research. Clearly, when cognitive distance is either too large or too small, this might be a hindrance to learning and innovation. *Boschma* [3.6] considers five types of proximity in the context of firms. We take the liberty of re-interpreting them in the context of collaborating research teams with a different disciplinary background.

3.15.1 Cognitive Proximity

Effective transfer of knowledge requires absorptive capacity to identify, interpret and exploit new knowledge. As a rule, researchers have a tendency to perform research in close proximity to their existing knowledge base. Yet, this sets constraints to further improvement. For each new step there exists a minimum level of knowledge under which teams are incapable of bridging the knowledge gap. That is, a team's cognitive base should be close enough to a collaborating team's knowledge base in order to successfully communicate, understand and process what the other team has to offer. Cognitive proximity facilitates communication, but too much cognitive proximity is detrimental to new learning and innovation. On the one hand, effective IDR needs a limited cognitive overlap, while on the other hand there must be sufficient cognitive overlap for the sake of communication. We note, though, that the act of collaborating itself increases cognitive proximity.

3.15.2 Organizational Proximity

Organizational practices are very relevant to the issue of IDR. Although a common knowledge and competence base is a prerequisite for bringing teams together, knowledge creation also depends on the capacity to coordinate the exchange of complementary pieces of knowledge owned by a variety of actors. Too much organizational proximity leads to a lack of flexibility, while too little organizational proximity leads to a lack of control and possible opportunistic behaviour (collaborators that mainly take, i. e., learn new techniques, and hardly share information and skills).

3.15.3 Social Proximity

Research is embedded in a social context. Social proximity leads to trust based on acquaintanceship, maybe even friendship and joint experiences. Social proximity reduces the risk of opportunistic behaviour. Too little social distance may weaken IDR due to an overload of trust (not all experiments or statistical analyses are redone or checked as far as possible by a member of the other team). Too little social proximity may lead to lack of trust and commitment to the joint project.

3.15.4 Institutional Proximity

Social proximity is understood in terms of individual scientists, while institutional proximity is related to the general institutional framework. Institutional proximity refers to norms and values of conduct. These may differ between departments and on a larger scale between universities and countries. These norms may be derived from legal laws and rules, but also from cultural norms and habits. A culture of shared trust is a basis for IDR. New experiences, leading to innovative research, are more easily transmitted in a sphere of cultural proximity and a common language. As such, institutional proximity provides stable underlying conditions for IDR. The cultural dimension of institutional proximity is related to geographical proximity, discussed next.

3.15.5 Geographical or Spatial Proximity

When co-workers are co-located it goes without saying that probably the other forms of proximity are also high. Yet, it may happen that two universities situated in the same city have quite different goals. Being geographically close may, however, lead to IDR that otherwise would not have taken place. Once teams are formed, geographical proximity may stimulate all other forms of proximity as face-to-face contact may easily increase. Moreover, geographic proximity may be a cost-reducing factor, as collaboration and face-to-face discussion may occur without incurring travel costs.

3.15.6 More on Cognitive Distance

Cognitive distance between research groups and members of an evaluation panel has been measured in [3.77, 91] based on researchers' publication portfolios. In these investigations, categories are either WoS subject categories or journals in which articles have been published. In the calculation of cognitive distance similarity between these categories or journals was taken into account.

Wang and *Sandström* [3.92] provide a methodology for the measurement of cognitive distance between researchers and study its role in peer review. Scientists holding the mainstream view of their field may not be able to give a fair review regarding new or alternative research. This is in particular the case for IDR. These authors developed a strategy that combines measuring research experience and content to obtain a measure for the cognitive distance between applicants and referees in a grant reviewing exercise. Note that grant reviewing panels are organized for a longer period of time, which is quite different from a one-off panel. In the first approach the authors use cited references as a researcher's knowledge base. Hence, the more references two researchers have in common the smaller their cognitive distance. This implies that the first method proposed in [3.92] is a form of author-bibliographic coupling. Based on the obtained author-reference matrix they calculate the Salton cosine measure between authors (practically: an applicant for a grant and a panel member). Cognitive distance is then defined as "one minus the cosine value" (where the cosine value was always positive in this application). It is not clear, however, what is a *good* outcome as values for a cognitive distance of zero or one are both undesirable. Similarly, in the second approach *Wang* and *Sandström* [3.92] analyze actual research contents, restricted to titles and abstracts and form an author-topic matrix. Again, they calculate a cosine measure between authors. In their discussion they mention that when it is desired to have industry representatives in a panel, their approach may not work, as these representatives typically have none or very few academic publications.

3.16 Identification of Interdisciplinary Ideas

The process of creating an innovative product starts with research activities based on a new idea. Decision makers would like to obtain estimations about the innovative potential of an idea before work on its realization starts. As estimating the interdisciplinarity of an idea can be done beforehand, interdisciplinarity is used as a proxy for innovative potential. The work of *Thorleuchter* and *Van den Poel* [3.93–95] aims at identifying the innovative potential of a new idea using texts (in theory one could also use auditory or visual information). In this way they hope to improve the performance of the innovation process. Innovation tends to occur between two or more technologies or two or more scientific fields that are not yet related. For this reason innovations can be said to be interdisciplinary products. An idea has two parts: a means (how) and a purpose (what). When an idea is represented as a text phrase the authors try to identify interdisciplinary ideas as cases where means and purpose are assigned to different not-yet-related technologies or scientific disciplines.

Their general approach is a form of text mining referred to as idea mining. In a first step, idea mining identifies means and ends (purposes) from textual information describing a problem. In a second step, it searches in new information sources for textual patterns that contain either a means or an end from the problem description and that also contain corresponding new means or new ends. This leads to a pattern of a *new* means leading to an *old* end; or vice versa. Idea mining usually identifies a large number of possible solutions. Next, possible solutions are classified and clustered. Relationships of these clusters to disciplines are determined to find out if an idea is of interdisciplinary nature or not. In this step one needs a description of fields or disciplines. *Thorleuchter* and *Van den Poel* [3.93] use the scope notes of the WoS subject categories for this purpose. As these scope notes are manually created, the authors claim that they are of good quality and up to date. To present the ideas in a comprehensible way to human experts, a text phrase is built for each idea. Then human experts evaluate the ideas shown to see if they can be put into practice. This approach is closely related to *Swanson*'s idea of literature-based discovery [3.96, 97].

3.17 Time Aspects

The fact that knowledge integration and diffusion themselves are dynamic processes has consequences for their measurement [3.23]. What is currently thought of as a highly interdisciplinary field is a point-in-time perception of how far apart the present constituent categories were at an earlier time, suggesting that a measure of the distance between two topics or between fields of science should be based on the analysis of large amounts of data over time. Moreover, interdisciplinary practices can be assumed to differ across disciplines. A medical researcher using basic statistics is not performing interdisciplinary work; yet, using statistics that requires a skilled statistician in the team, is. This shows the influence of basic education in a field (what is included in basic teachings and what is not) on the notion of IDR.

For these reasons, IDR can be considered as part of an evolution, rather than a state, thereby requiring that researchers must portray its development over time. In fact, given the consensus that a central concept of IDR is a process of knowledge integration, this in itself suggests a dynamic process leading from less to more integration. Yet, established fields may have subfields that start diverging [3.98]. Stated otherwise, on the micro level one might see knowledge integration but on the macro level of fields or subfields one may observe divergence. While an existing field might be characterized by a static IDR measure as *highly integrated*, in the process of evolving to its highly integrated state, it may have lost much of its novelty and breakthrough value. Hence being highly integrated is not a desirable state, as already mentioned in Sect. 3.10 (high diversity and high coherence). Insofar as current IDR measurement is often static, using a finite time window, this most certainly is not an optimal state of affairs.

3.18 Limitations of Existing Approaches

Existing databases have a number of limitations that raise questions about their usefulness in developing bibliometric-based measures of interdisciplinary research. Funding agencies often seek to support research that addresses complex scientific and social problems. They claim that doing so requires the integrated contributions of varied fields of knowledge. This clearly implies that the humanities and social sciences must play an important role in such investigations. Yet, this aspect has often been neglected when applying ideas, concepts and measuring techniques conceived for the natural sciences and the life sciences. Emphasis on products from Clarivate Analytics, such as the JCR, or Elsevier's Scopus, may well hinder developments in the study of IDR. Reliable data to assess developments in the SS&H do not exist on an international level because the international databases only partially cover books, book chapters and especially regional non-English journals. This implies that a full-scope IDR study can only be done when incorporating regional databases. *Wagner* et al. [3.2] write that, given the limits of bibliometrics, we may be missing the most socially relevant IDR interactions because of limitations in the current contents of bibliographic databases, or, stated otherwise, because researchers who study IDR take the easy way-out and only use these databases in their studies.

3.19 An Example Within the Rafols–Meyer Framework

In this section we provide a simple example of a study on interdisciplinarity, applying the methodology suggested in [3.8]. Fourteen articles in the field of synthetic biology [3.99–112], a field in which we have some prior experience [3.113], were chosen as illustration. All these articles were published in the year 2000.

There is one important difference with [3.8] though: we use the true diversity measure derived from the Rao–Stirling measure instead of the classical one [3.25, 114, 115]. Concretely, as diversity measure we use Z, defined as

$$Z = \frac{1}{1-D}\,, \tag{3.8}$$

where D is the Rao–Stirling measure. Coherence is calculated as mean linkage strength for the network of normalized (using Salton's cosine) bibliographic coupling strength [3.8], between references included in the Web of Science.

Concretely, if an article P has ten references, among which five are items not included in the WoS, e. g., articles in journals not covered in the WoS, books, an unpublished doctoral thesis, then these items are discarded. Taken over the 14 articles studied as an illustration, the percentage of discarded items is about 30%.

The remaining normalized bibliographic coupling matrix $\mathbf{M}_\mathrm{P}$ may be of the form shown below, where the diagonal values play no role and are indicated by a hyphen

$$\mathbf{M}_\mathrm{P} = \begin{pmatrix} - & 0.25 & 0.05 & 0.10 & 0.00 \\ 0.25 & - & 0.30 & 0.02 & 0.00 \\ 0.05 & 0.30 & - & 0.01 & 0.00 \\ 0.10 & 0.02 & 0.01 & - & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & - \end{pmatrix},$$

where the normalized bibliographic coupling values between each two references were calculated based on the Salton cosine measure as follows: if *shared* refers to the number of shared references between Ref 1 and Ref 2 and $T1$, resp. $T2$, refer to the number of references in Ref 1, resp. Ref 2, then the Salton cosine measure is

$$\frac{\text{shared}}{\sqrt{T1 * T2}}\,.$$

Here, the *number of references in Ref k* $k = 1, 2$, refers only to those included in the WoS.

The mean linkage strength is the average of all non-diagonal values. In this example, the coherence value for article B is

$$\begin{aligned}\text{Coherence} &= \frac{\text{sum of all elements in } \mathbf{M}_\mathrm{P}}{20} \\ &= 0.073\,.\end{aligned}$$

The final results in terms of diversity and coherence for each article in the example set are shown in Fig. 3.2.

Figure 3.3 illustrates the differences in network structure associated with increasing values for coherence, for 4 articles in the sample.

Figure 3.4 illustrates the disciplinary mix of article E's reference set by locating the WoS subject categories where references were published. Note that the same reference may be assigned to different subject categories.

These examples illustrate not only the methodology used to study interdisciplinarity, but also the interaction of fields leading to studies in synthetic biology.

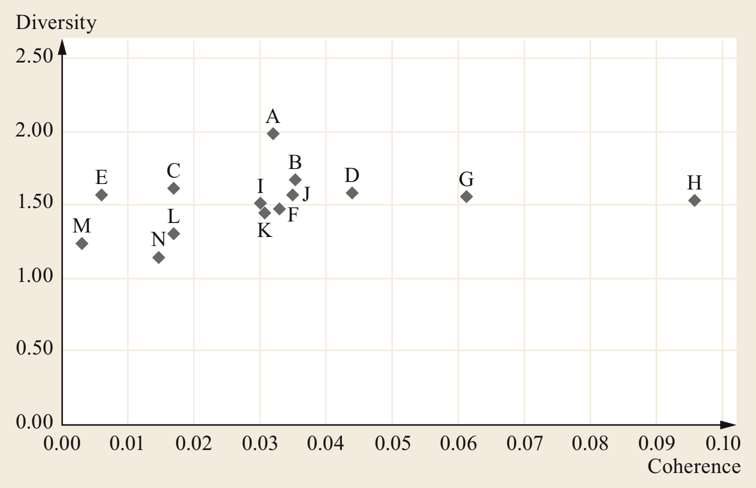


Fig. 3.2 Diversity and coherence for 14 articles in the field of synthetic biology: Coherence on the horizontal axis; diversity (Z) on the vertical axis. *Letter symbols A–N* refer to references [3.99–112], in that order

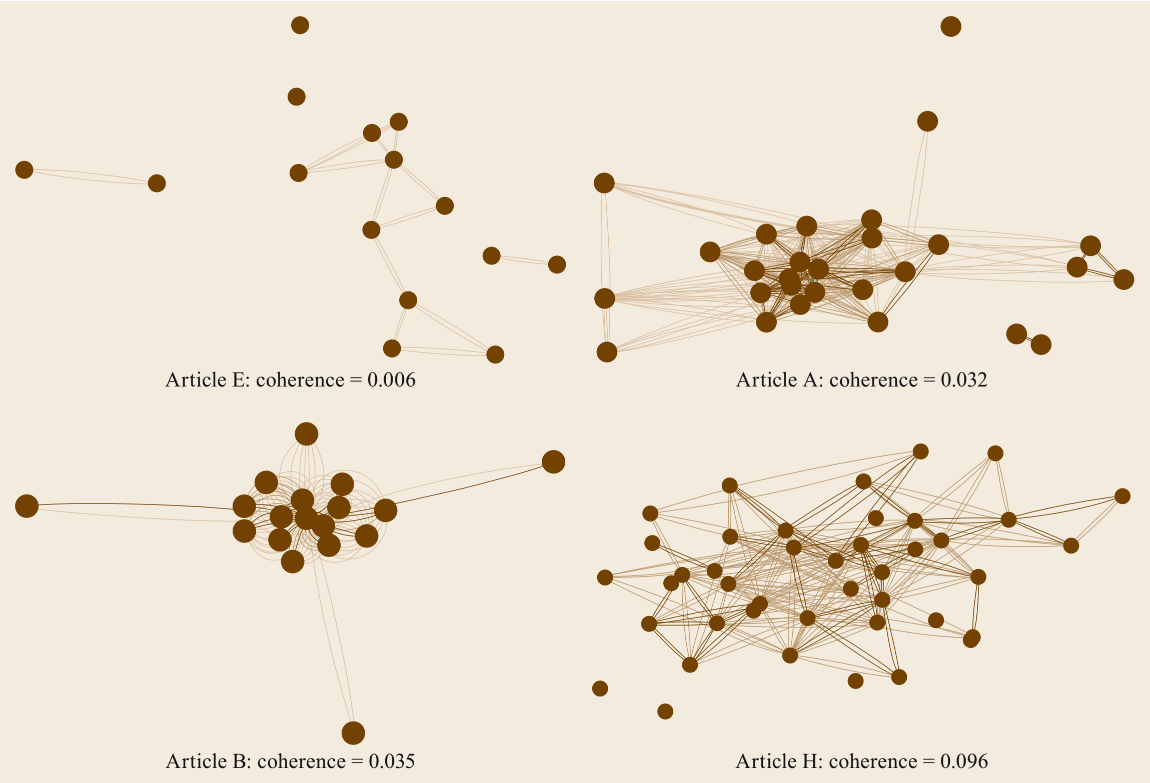


Fig. 3.3 Bibliographic coupling networks for the reference set of various articles. The figures are ordered from lower to higher network coherence (from *top left* to *bottom right*); *thicker lines* indicate greater similarity. Also here, *letter symbols A–N* refer to references [3.99–112], in that order

Fig. 3.4 Distribution of WoS subject categories for the reference set of article E [3.103] in bibliographic coupling networks. *Black nodes* indicates the references published in a given subject category (shown under each tile). *Brown nodes* indicate the references published in other subject categories (different from the given one as indicated by the *black nodes*)

3.20 Conclusions and Suggestions for Further Research

Following the definition of IDR according to the National Academies of Sciences of the USA, the essence of IDR is knowledge integration. Yet, besides integration of knowledge originating in different disciplines, we pointed out that knowledge may also originate from different research traditions, different regions and different ages.

As there does not exist a definition of the term *discipline* accepted by the majority of scientists, one must measure ID based on some operationalization. Often the choice goes to WoS categories, although it is well-known that such a choice is far from optimal.

We discussed inputs, outcomes and the process of IDR itself. IDR normally involves collaboration by scientists with a different background. This leads to different perspectives on the research that is performed and avoids lock-ins. We further discussed, be it briefly, some costs associated with doing IDR, such as lack of appreciation and being less funded.

On the one hand, quantitative measurement of IDR often implies the use of references (the citing side of the article or group of articles on which the measurement is performed). Diffusion studies, on the other hand, consider the cited side. *Rafols* and *Meyer* [3.8] used a top-down approach for measuring diversity among references and a bottom-up approach for coherence.

Once a measure of ID has been determined, one may answer the question if more ID research receives more citations. Investigations trying to answer this question indicated that the answer is no. The most-cited publications are those with a clear disciplinary focus, but that nevertheless give small proportions of references to many proximal disciplinary categories.

Evaluation of IDR itself has been described following [3.7] while different types of proximities have been described following [3.6].

This review shows that knowledge integration and diffusion are essentially dynamical processes in scientific evolution. Involving different disciplines, different regions, and different countries, these processes constitute the main motors of creativity and innovativeness, leading to new scientific structures of science, but based on the traditional fields as foundation.

Measuring or studying interdisciplinarity and, more generally, knowledge integration and diffusion, are always subjective actions, depending on several choices. Indeed, as long as there is no generally accepted definition of a discipline, ID cannot be but an artificial construct, open to diverse interpretations.

Following [3.2] we saw that assessment of interdisciplinary research inputs, processes, outputs and outcomes is still a work in progress. There are, for instance, few studies that link inputs to outputs, let alone outcomes.

The authors of [3.47] write that scrutiny reveals the limitations of an obsession with performance (especially as measured through evaluation assessments such as the Research Excellence Framework in the UK). Structural diversity is a necessary complement to research excellence. At the university level, a department focusing on high-impact journals may not be in the best position to address interdisciplinary policy challenges. At the country level, the research base with the highest average citation impact does not necessarily have the best long-term portfolio. They further write that because of humans' uncertainty about the future we need an agile and responsive research base, leading to the question of finding a measure of *agility*. Probably such a research base must, much more than now, include the social sciences and humanities.

When multiple studies have determined the degree of ID, the next step can be set with confidence; that is, studying whether investigations with a high degree of ID are the ones that have the most impact, leading to the real *novel* results aimed for by governments and funders. Further research will have to determine the links between *interdisciplinary research* and *novel research*. During this step experts from the targeted scientific/technology fields must play a decisive role.

When it comes to studying how IDR reaches the general public, alternative metrics may constitute an obvious point of departure.

In view of *Stirling*'s work [3.69] and that of others [3.8, 116] we come to the conclusion that, ideally, measures of diversity:

(a) Satisfy reasonable axioms, as pointed out in [3.69] and studied in detail by *Leinster* and *Cobbold* [3.114].
(b) Are parametrized to produce diversity profiles [3.114]; concretely, next to Z as used in the example, one must also consider the family

$$Z^q = \left(\sum_{i=1}^{N} p_i \left(\sum_{j=1}^{N} s_{ij} p_j \right)^{q-1} \right)^{\frac{1}{1-q}} , \tag{3.9}$$

where q is a parameter ranging from zero to infinity (the cases $q = 1$ and $q = \infty$ are obtained as limits); $\mathbf{S} = (s_{ij})_{ij}$ is a similarity matrix; the corresponding disparity matrix $\mathbf{D}$ is obtained as $d_{ij} = 1 - s_{ij}$; Z is the case $q = 2$.
(c) Come with an effective number to simplify interpretation: this means that the result can be interpreted

for a community of equally abundant, totally dissimilar species [3.65, 114, 117].
(d) Have (unbiased) estimators to apply to real data.

The first three points fall within the framework of this chapter, while we leave the discussion of the statistical point (d), to specialists such as [3.116].

Which measures are most appropriate for addressing particular questions concerning the interdisciplinary content of research output, at what levels of aggregation and with what degree of validity and reliability, is still a huge research question and a topic of further discussion [3.2].

Structural diversity should be placed into an analytical framework that provides acceptable general measures, relevant to policy and meaningful to stakeholders. The results of such measures would be compared with models of desirable outcomes. Only then is it possible to show where diversity has been of value and track where this valuable diversity might be lost, and hence take steps to safeguard these strengths [3.47].

Most existing approaches to measure interdisciplinarity are just drawing up the *shape* of an interdisciplinary field, and are less interested in the effect generated by interdisciplinary work. This important perspective is suggested for future studies.

Although investment must support current policy priorities, it must be accepted that without curiosity-driven research identified by researchers themselves, one would soon be mining worn-out seams [3.47]. Consequently, we conclude that mono-disciplinary and interdisciplinary work and projects both have their place in producing research that has an impact on other fields and on society.

In this review we did not discuss specific case studies published in the literature, leaving that to another occasion. Yet, we do draw the reader's attention to [3.118]. This article discusses IDR which aims at integrating the social and natural sciences, a particularly ambitious endeavor. *Roessner* et al. [3.119] investigated how features of IDR are accurately reflected in bibliometric measures of scholarly publications, concretely the Rao–Stirling integration score, over time, focusing on the entire portfolio of a well-known researcher who was active in different fields. In their mapping study *Boyack* et al. [3.71] found that biochemistry was the most interdisciplinary discipline in science.

Acknowledgments. Lin Zhang acknowledges support from the National Natural Science Foundation of China Grant 71573085. Xiaojun Hu and Ronald Rousseau acknowledge support from the National Natural Science Foundation of China Grant 71573225. We further thank Loet Leydesdorff and Raf Guns for useful comments.

References

3.1 M. Dogan, R. Pahre: *Creative Marginality: Innovation at the Intersections of Social Sciences* (Westview, Boulder 1990)
3.2 C.S. Wagner, J.D. Roessner, K. Bobb, J.T. Klein, K.W. Boyack, J. Keyton, I. Rafols, K. Börner: Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature, J. Informetr. **5**(1), 14–26 (2011)
3.3 National Science Board: *Science and Engineering Indicators 2016* (National Science Foundation, Arlington 2016), NSB-2016-1
3.4 ERC: ERC Frontier Research Grants Information for Applicants to the Starting and Consolidator Grant 2016 Calls, http://ec.europa.eu/research/participants/data/ref/h2020/other/guides_for_applicants/h2020-guide16-erc-stg-cog_en.pdf (2015)
3.5 J. Adams, T. Loach, M. Szomszor: *The Diversity of UK Research and Knowledge. Digital Research Reports* (Digital Science, London 2015)
3.6 R.A. Boschma: Proximity and innovation: A critical assessment, Reg. Stud. **39**(1), 61–74 (2005)
3.7 J.T. Klein: Evaluation of interdisciplinary and transdisciplinary research, Am. J. Prev. Med. **35**(2S), S116–S123 (2008)
3.8 I. Rafols, M. Meyer: Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience, Scientometrics **82**(2), 263–287 (2010)
3.9 I. Rafols: Knowledge integration and diffusion: Measures and mapping of diversity and coherence. In: *Measuring Scholarly Impact*, ed. by Y. Ding, R. Rousseau, D. Wolfram (Springer, Cham 2014) pp. 169–190
3.10 National Academies: *Facilitating Interdisciplinary Research* (National Academies, Washington DC 2005)
3.11 OECD: *A Framework for Biotechnology Statistics* (OECD, Paris 2005)
3.12 N. Carayol, T.U.N. Thi: Why do academic scientists engage in interdisciplinary research?, Res. Eval. **14**(1), 70–79 (2005)
3.13 J. Molas-Gallart, I. Rafols, P. Tang: On the relationship between interdisciplinarity and impact: Different modalities of interdisciplinarity lead to different types of impact, J. Sci. Policy Res. Manag. **29**(2/3), 69–89 (2014)
3.14 Y.X. Liu, I. Rafols, R. Rousseau: A framework for knowledge integration and diffusion, J. Doc. **68**(1), 31–44 (2012)

3.15 V.D. Blondel, J.L. Guillaume, R. Lambiotte, E. Lefebvre: Fast unfolding of communities in large networks, J. Stat. Mech. Theory Exp. P10008 (2008)
3.16 M. Rosvall, C.T. Bergstrom: Maps of random walks on complex networks reveal community structure, Proc. Natl. Acad. Sci. USA **105**(4), 1118–1123 (2008)
3.17 I. Rafols, L. Leydesdorff: Content-based and algorithmic classifications of journals: Perspectives on the dynamics of scientific communication and indexer effects, J. Am. Soc. Inf. Sci. Technol. **60**(9), 1823–1835 (2009)
3.18 W. Glänzel, A. Schubert: A new classification scheme of science fields and subfields designed for scientometric evaluation purposes, Scientometrics **56**(3), 357–367 (2003)
3.19 L. Leydesdorff, I. Rafols: A global map of science based on the ISI subject categories, J. Am. Soc. Inf. Sci. Technol. **60**(2), 348–362 (2009)
3.20 A.L. Porter, A.S. Cohen, J.D. Roessner, M. Perreault: Measuring researcher interdisciplinarity, Scientometrics **72**(1), 117–147 (2007)
3.21 J. Adams, T. Loach, M. Szomszor: *Interdisciplinary Research: Methodologies for Identification and Assessment. Digital Research Reports* (Digital Science, London 2016)
3.22 L. Leydesdorff, L. Bornmann: The operationalization of "fields" as WoS subject categories (WCs) in evaluative bibliometrics: The cases of "library and information science" and "science and technology studies", J. Assoc. Inf. Sci. Technol. **67**(3), 707–714 (2016)
3.23 Y.X. Liu, R. Rousseau: Knowledge diffusion through publications and citations: A case study using ESI-fields as unit of diffusion, J. Am. Soc. Inf. Sci. Technol. **61**(2), 340–351 (2010)
3.24 L. Bromham, R. Dinnage, X. Hua: Interdisciplinary research has consistently lower funding success, Nature **534**(7609), 684–687 (2016)
3.25 L. Zhang, R. Rousseau, W. Glänzel: Diversity of references as an indicator of the interdisciplinarity of journals: Taking similarity between subject fields into account, J. Assoc. Inf. Sci. Technol. **67**(5), 1257–1265 (2016)
3.26 L. Leydesdorff, D. Rotolo, I. Rafols: Bibliometric perspectives on medical innovation using the Medical Subject Headings of PubMed, J. Am. Soc. Inf. Sci. Technol. **63**(11), 2239–2253 (2012)
3.27 A. Skupin, J.R. Biberstine, K. Börner: Visualizing the topical structure of the medical sciences: A self-organizing map approach, PLoS One **8**(3), e58779 (2013)
3.28 L. Smolinsky, A. Lercher: Citation rates in mathematics: A study of variation by subdiscipline, Scientometrics **91**(3), 911–924 (2012)
3.29 H. Mihaljević-Brandt, L. Santamaría, M. Tullney: The effect of gender in the publication patterns in mathematics, PLoS One **11**(10), e0165367 (2016)
3.30 S. Bhattacharya, S.P. Singh, P. Sudhakar: Tracking changes in research priorities in physics: A macro level analysis, Scientometrics **40**(1), 57–82 (1997)
3.31 Z.S. Shen, L.Y. Yang, J.S. Pei, M.H. Li, C.S. Wu, J.Z. Bao, T. Wei, Z.R. Di, R. Rousseau, J.S. Wu: Interrelations among scientific fields and their relative influences revealed by an input-output analysis, J. Informetr. **10**(1), 82–97 (2016)
3.32 X.J. Hu, R. Rousseau: A simple approach to describe a company's innovative activities and their technological breadth, Scientometrics **102**(2), 1401–1411 (2015)
3.33 X.J. Hu, R. Rousseau: Gauging a firm's innovative performance using an integrated structural index for patents, J. Data Inf. Sci. **1**(1), 6–27 (2016)
3.34 J. Adams, L. Jackson, S. Marshall: *Bibliometric Analysis of Interdisciplinary Research*, Report to the Higher Education Funding Council for England (Evidence, Leeds 2007)
3.35 A. Yegros-Yegros, I. Rafols, P. D'Este: Does interdisciplinary research lead to higher citation impact? The different effect of proximal and distal interdisciplinarity, PLoS One **10**(8), e0135095 (2015)
3.36 D.E. Stokes: *Pasteur's Quadrant–Basic Science and Technological Innovation* (Brookings Institution, Washington DC 1997)
3.37 C. Haythornthwaite: Learning and knowledge networks in interdisciplinary collaborations, J. Am. Soc. Inf. Sci. Technol. **57**(8), 1079–1092 (2006)
3.38 J.S. Katz, B.R. Martin: What is research collaboration?, Res. Policy **26**(1), 1–18 (1997)
3.39 G. Mallard, M. Lamont, J. Guetzkow: Fairness as appropriateness: Negotiating epistemological differences in peer review, Sci. Technol. Hum. Values **34**(5), 573–606 (2009)
3.40 S. De Rijcke, P.F. Wouters, A.D. Rushforth, T.P. Franssen, B. Hammarfelt: Evaluation practices and effects of indicator use—A literature review, Res. Eval. **25**(2), 161–169 (2016)
3.41 E. Leahey, C.M. Beckman, T.L. Stanko: Prominent but less productive: The impact of interdisciplinarity on scientists' research, Adm. Sci. Q. **62**(1), 105–139 (2017)
3.42 D. Stokols, J. Fuqua, J. Gress, R. Harvey, K. Phillips, L. Baezconde-Garbanati, J. Unger, P. Palmer, M.A. Clark, S.M. Colby, M. Morgan, W. Trochim: Evaluating transdisciplinary science, Nicotine Tob. Res. **5**(1), S21–S39 (2003)
3.43 B. Hanscom, H.E. Janes, P.D. Guarino, Y. Huang, E.R. Brown, Y.Q. Chen, S.M. Hammer, P.B. Gilbert, D.J. Donnell: Brief Report: Preventing HIV-1 infection in women using oral preexposure prophylaxis: A meta-analysis of current evidence, J. Acquir. Immune Defic. Syndr. **73**(5), 606–608 (2016)
3.44 L. Sanz-Menéndez, M. Bordons, M.A. Zulueta: Interdisciplinarity as a multidimensional concept: Its measure in three different research areas, Res. Eval. **10**(1), 47–58 (2001)
3.45 L.C. Smith: Citation analysis, Libr. Trends **30**(Summer), 83–106 (1981)
3.46 G. Lewison: Beyond outputs: New measures of biomedical research impact, ASLIB Proceedings **55**(1/2), 32–42 (2003)
3.47 Digital Science: *The Value of Structural Diversity: Assessing Diversity for A Sustainable Research Base* (Digital Science, London 2015)

3.48 R. May: The scientific wealth of nations, Science **275**(5301), 793–796 (1997)
3.49 L. Cassi, W. Mescheba, E. de Turckheim: How to evaluate the degree of interdisciplinarity of an institution?, Scientometrics **101**(3), 1871–1895 (2014)
3.50 M. Zitt: Facing diversity of science: A challenge for bibliometric indicators, Measurement **3**(1), 38–49 (2005)
3.51 L.G. Nichols: A topic model approach to measuring interdisciplinarity at the National Science Foundation, Scientometrics **100**(3), 741–754 (2014)
3.52 E.J. Rinia, T.N. van Leeuwen, A.F.J. van Raan: Impact measures of interdisciplinary research in physics, Scientometrics **53**(2), 241–248 (2002)
3.53 V. Larivière, Y. Gingras: On the relationship between interdisciplinarity and scientific impact, J. Am. Soc. Inf. Sci. Technol. **61**(1), 126–131 (2010)
3.54 E. Garfield, M.V. Malin, H. Small: Citation data as science indicators. In: *Toward a Metric of Science: The Advent of Science Indicators*, ed. by Y. Elkana, J. Lederberg, R.K. Merton, A. Thackray, H. Zuckerman (Wiley, New York 1978) pp. 179–207
3.55 M. Zitt, S. Ramanana-Rahary, E. Bassecoulard: Relativity of citation performance and excellence measures; from cross-field to cross-scale effects of field-normalisation, Scientometrics **63**(2), 373–401 (2005)
3.56 L. Leydesdorff: Betweenness centrality as an indicator of the interdisciplinarity of scientific journals, J. Assoc. Inf. Sci. Technol. **58**(9), 1303–1319 (2007)
3.57 L. Leydesdorff: Mapping interdisciplinarity at the interfaces between the science citation index and the social science citation index, Scientometrics **71**(3), 391–405 (2007)
3.58 G. Abramo, C.A. D'Angelo, F. Di Costa: Identifying interdisciplinarity through the disciplinary classification of coauthors of scientific publications, J. Am. Soc. Inf. Sci. Technol. **63**(11), 2206–2222 (2012)
3.59 A.F.J. van Raan: Scaling rules in the science system: Influence of field-specific citation characteristics on the impact of research groups, J. Assoc. Inf. Sci. Technol. **59**(4), 565–576 (2008)
3.60 V. Batagelj, A. Mrvar: Pajek–Program for large network analysis, Connections **21**(2), 47–57 (1998)
3.61 N.J. van Eck, L. Waltman: VOS: A new method for visualizing similarities between objects. In: *Advances in Data Analysis: Proceedings of the 30th Annual Conference of the German Classification Society*, ed. by H.-J. Lenz, R. Decker (Springer, Heidelberg 2007) pp. 299–306
3.62 N.J. van Eck, L. Waltman: Software survey: VOSviewer, A computer program for bibliometric mapping, Scientometrics **84**(2), 523–538 (2010)
3.63 P. van den Besselaar, L. Leydesdorff: Mapping change in scientific specialties: A scientometric reconstruction of the development of artificial intelligence, J. Am. Soc. Inf. Sci. **47**(6), 415–436 (1996)
3.64 P. van den Besselaar, G. Heimeriks: Disciplinary, multidisciplinary, interdisciplinary: Concepts and indicators. In: *Proc. 8th Int. Conf. Scientometr. Informetr., Sidney* (2001) pp. 705–716
3.65 L. Jost: Entropy and diversity, Oikos **113**(2), 363–375 (2006)
3.66 F. Morillo, M. Bordons, I. Gomez: An approach to interdisciplinarity through bibliometric indicators, Scientometrics **51**(1), 203–222 (2001)
3.67 F. Morillo, M. Bordons, I. Gomez: Interdisciplinarity in science: A tentative typology of disciplines and research areas, J. Am. Soc. Inf. Sci. Technol. **54**(13), 1237–1249 (2003)
3.68 A. Porter, I. Rafols: Is science becoming more interdisciplinary? Measuring and mapping six research fields over time, Scientometrics **81**(3), 719–745 (2009)
3.69 A. Stirling: A general framework for analysing diversity in science, technology and society, J. Royal Soc. Interface **4**(15), 707–719 (2007)
3.70 K.W. Boyack, R. Klavans, H. Small, L. Ungar: Characterizing the emergence of two nanotechnology topics using a contemporaneous global micro-model of science, J. Eng. Technol. Manag. **32**, 147–159 (2014)
3.71 K.W. Boyack, R. Klavans, K. Börner: Mapping the backbone of science, Scientometrics **64**(3), 351–374 (2005)
3.72 F. Moya-Anegón, B. Vargas-Quesada, Z. Chinchilla-Rodríguez, E. Corera-Álvarez, F.J. Munoz-Fernández, V. Herrero-Solana: Visualizing the marrow of science, J. Am. Soc. Inf. Sci. Technol. **58**(14), 2167–2179 (2007)
3.73 R. Klavans, K. Boyack: Toward a consensus map of science, J. Am. Soc. Inf. Sci. Technol. **60**(3), 455–476 (2009)
3.74 I. Rafols, A.L. Porter, L. Leydesdorff: Science overlay maps: A new tool for research policy and library management, J. Am. Soc. Inf. Sci. Technol. **61**(9), 1871–1887 (2010)
3.75 L. Leydesdorff, S. Carley, I. Rafols: Global maps of science based on the new Web-of-Science categories, Scientometrics **94**(2), 589–593 (2013)
3.76 L. Leydesdorff, L. Bornmann, P. Zhou: Construction of a pragmatic base line for journal classifications and maps based on aggregated journal-journal citation relations, J. Informetr. **10**(4), 902–918 (2016)
3.77 R. Rousseau, R. Guns, A.I.M.J. Rahman, T. Engels: Measuring cognitive distance between publication portfolios, J. Informetr. **11**(2), 583–594 (2017)
3.78 F.N. Silva, F.A. Rodrigues, O.N. Oliveira Jr, L.F. da Costa: Quantifying the interdisciplinarity of scientific journals and fields, J. Informetr. **7**(2), 469–477 (2013)
3.79 C. Chen, Y. Chen, M. Horowitz, H.Y. Hou, Z.Y. Liu, D. Pellegrino: Towards an explanatory and computational theory of scientific discovery, J. Informetr. **3**(3), 191–209 (2009)
3.80 T. Luukkonen, M. Nedeva: Towards understanding integration in research and research policy, Res. Policy **39**(5), 674–686 (2010)
3.81 R. Boschma, P.-A. Balland, M. de Vaan: The formation of economic networks: A proximity approach. In: *Regional Development and Proximity Relations*, ed. by A. Torre, F. Wallet (Edward Elgar, Cheltenham 2014) pp. 243–266

3.82 Y.X. Liu, R. Rousseau, R. Guns: A layered framework to study collaboration as a form of knowledge sharing and diffusion, J. Informetr. **7**(3), 651–664 (2013)
3.83 I. Rafols, L. Leydesdorff, A. O'Hare, P. Nightingale, A. Stirling: How journal rankings can suppress interdisciplinary research: A comparison between innovation studies and business and management, Res. Policy **41**(7), 1262–1282 (2012)
3.84 V. Boix-Mansilla, I. Feller, H. Gardner: Quality assessment in interdisciplinary research and education, Res. Eval. **15**(1), 69–74 (2006)
3.85 B.H. Jin, R. Rousseau, X.X. Sun: Key labs and open labs in the Chinese scientific research system: Qualitative and quantitative evaluation indicators, Res. Eval. **14**(2), 103–109 (2005)
3.86 L.R. Meagher, U. Martin: Slightly dirty maths: The richly textured mechanisms of impact, Res. Eval. **26**(1), 15–27 (2017)
3.87 N.J. van Eck, L. Waltman, A.F.J. van Raan, R.J.M. Klautz, W.C. Peul: Citation analysis may severely underestimate the impact of clinical research as compared to basic research, PLoS One **8**(4), e62395 (2013)
3.88 S.J. Chen, C. Arsenault, V. Larivière: Are top-cited papers more interdisciplinary?, J. Informetr. **9**(4), 1034–1046 (2015)
3.89 J. Wang, B. Thijs, W. Glänzel: Interdisciplinarity and impact: Distinct effects of variety, balance, and disparity, PLoS One **10**(5), e0127298 (2015)
3.90 A. Broström, M. McKelvey: *Knowledge Transfer at the Science-Policy Interface: How Cognitive Distance and the Degree of Expert Autonomy Shapes the Outcome*, Working Paper Series in Economics and Institutions of Innovation from Royal Institute of Technology, Vol. 441 (CESIS: Centre of Excellence for Science and Innovation Studies, Stockholm 2016)
3.91 A.I.M.J. Rahman, R. Guns, L. Leydesdorff, T. Engels: Measuring the match between evaluators and evaluees: Cognitive distances between panel members and research groups at the journal level, Scientometrics **109**(3), 1639–1663 (2016)
3.92 Q. Wang, U. Sandström: Defining the role of cognitive distance in the peer review process with an explorative study of a grant scheme in infection biology, Res. Eval. **24**(3), 271–281 (2015)
3.93 D. Thorleuchter, D. Van den Poel: Identification of interdisciplinary ideas, Inf. Process. Manag. **52**(6), 1074–1085 (2016)
3.94 D. Thorleuchter, D. Van den Poel, A. Prinzie: Mining ideas from textual information, Expert Syst. Appl. **37**(10), 7182–7188 (2010)
3.95 D. Thorleuchter, D. Van den Poel, A. Prinzie: Mining innovative ideas to support new product research and development. In: *Classification as a Tool for Research*, ed. by H. Locarek-Junge, C. Weihs (Springer, Berlin 2010) pp. 587–594
3.96 D.R. Swanson: Undiscovered public knowledge, Libr. Q. **56**(2), 103–118 (1986)
3.97 D.R. Swanson: Fish oil, Raynaud's syndrome and undiscovered public knowledge, Perspect. Biol. Med. **30**(1), 7–18 (1986)
3.98 W. Glänzel, B. Thijs: Using 'core documents' for detecting and labelling new emerging topics, Scientometrics **91**(2), 399–416 (2012)
3.99 A. Kondo, S. Isaji, Y. Nishimura, T. Tanaka: Transcriptional and post-transcriptional regulation of monocyte chemoattractant protein-3 gene expression in human endothelial cells by phorbol ester and cAMP signalling, Immunology **99**(4), 561–568 (2000)
3.100 Y. Aihara, H. Mashima, H. Onda, S. Hisano, H. Kasuya, T. Hori, S. Yamada, H. Tomura, Y. Yamada, I. Inoue, I. Kojima, J. Takeda: Molecular cloning of a novel brain-type Na^+-dependent inorganic phosphate cotransporter, J. Neurochem. **74**(6), 2622–2625 (2000)
3.101 A.J. Butte, P. Tamayo, D. Slonim, T.R. Golub, I.S. Kohane: Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks, Proc. Natl. Acad. Sci. USA **97**(22), 12182–12186 (2000)
3.102 L. Dumoutier, J. Louahed, J.C. Renauld: Cloning and characterization of IL-10-related T cell-derived inducible factor (IL-TIF), a novel cytokine structurally related to IL-10 and inducible by IL-9, J. Immunol. **164**(4), 1814–1819 (2000)
3.103 K. Madin, T. Sawasaki, T. Ogasawara, Y. Endo: A highly efficient and robust cell-free protein synthesis system prepared from wheat embryos: Plants apparently contain a suicide system directed at ribosomes, Proc. Natl. Acad. Sci. USA **97**(2), 559–564 (2000)
3.104 H. Mattoussi, J.M. Mauro, E.R. Goldman, G.P. Anderson, V.C. Sundar, F.V. Mikulec, M.G. Bawendi: Self-assembly of CdSe-ZnS quantum dot bioconjugates using an engineered recombinant protein, J. Am. Chem. Soc. **122**(49), 12142–12150 (2000)
3.105 H.B. Li, A.F. Oberhauser, S.B. Fowler, J. Clarke, J.M. Fernandez: Atomic force microscopy reveals the mechanical design of a modular protein, Proc. Natl. Acad. Sci. USA **97**(12), 6527–6531 (2000)
3.106 S. Pan, R. Tsuruta, E.S. Masuda, R. Imamura, F. Bazan, K. Arai, N. Arai, S. Miyatake: NFATz: A novel rel similarity domain containing protein, Biochem. Biophys. Res. Commun. **272**(3), 765–776 (2000)
3.107 M. Miyazu, T. Tanimura, M. Sokabe: Molecular cloning and characterization of a putative cyclic nucleotide-gated channel from Drosophila melanogaster, Insect Mol. Biol. **9**(3), 283–292 (2000)
3.108 P.L. Ward, B. Taddeo, N.S. Markovitz, B. Roizman: Identification of a novel expressed open reading frame situated between genes U(L)20 and U(L)21 of the herpes simplex virus 1 genome, Virology **266**(2), 275–285 (2000)
3.109 S. Romer, P.D. Fraser, J.W. Kiano, C.A. Shipton, N. Misawa, W. Schuch, P.M. Bramley: Elevation of the provitamin A content of transgenic tomato plants, Nat. Biotechnol. **18**(6), 666–669 (2000)
3.110 T.O. Nielsen, N.H. Cossons, M. Zannis-Hadjopoulos, G.B. Price: Circular YAC vectors containing short

mammalian origin sequences are maintained under selection as HeLa episomes, J. Cell. Biochem. **76**(4), 674–685 (2000)
3.111 T.S. Gardner, C.R. Cantor, J.J. Collins: Construction of a genetic toggle switch in Escherichia coli, Nature **403**(6767), 339–342 (2000)
3.112 B. Ren, F. Robert, J.J. Wyrick, O. Aparicio, E.G. Jennings, I. Simon, J. Zeitlinger, J. Schreiber, N. Hannett, E. Kanin, T.L. Volkert, C.J. Wilson, S.P. Bell, R.A. Young: Genome-wide location and function of DNA binding proteins, Science **290**(5500), 2306–2309 (2000)
3.113 X.J. Hu, R. Rousseau: From a word to a world: The current situation in the interdisciplinary field of synthetic biology, PeerJ **3**, e728 (2015)
3.114 T. Leinster, C.A. Cobbold: Measuring diversity: The importance of species similarity, Ecology **93**(3), 477–489 (2012)
3.115 A.M. Mugabushaka, A. Kyriakou, T. Papazoglou: Bibliometric indicators of interdisciplinarity: The potential of the Leinster–Cobbold diversity indices to study disciplinary diversity, Scientometrics **107**(2), 593–607 (2016)
3.116 M. Grabchak, E. Marcon, G. Lang, Z.Y. Zhang: The generalized Simpson's entropy is a measure of biodiversity, PLoS One **12**(3), e0173305 (2017)
3.117 L. Jost: Mismeasuring biological diversity: Response to Hoffmann and Hoffmann (2008), Ecol. Econ. **68**(4), 925–928 (2009)
3.118 J. Garner, A. Porter, M. Borrego, E. Tran, R. Teutonico: Facilitating social and natural science cross-disciplinarity: Assessing the human and social dynamics program, Res. Eval. **22**(2), 134–144 (2013)
3.119 D. Roessner, A.L. Porter, N.J. Nersessian, S. Carley: Validating indicators of interdisciplinarity: Linking bibliometric measures to studies of engineering research labs, Scientometrics **94**(2), 439–468 (2013)

Ronald Rousseau
ECOOM
KU Leuven
Leuven, Belgium
ronald.rousseau@kuleuven.be

Ronald Rousseau is a Researcher at KU Leuven and at the University of Antwerp. His main interest is in mathematical aspects of citation analysis. He obtained a PhD in Mathematics (1977) and a Habilitations degree from KU Leuven (1983), and a PhD in Library and Information Science from Antwerp University (1992). From 2007 to 2015 he was the President of the International Society for Scientometrics and Informetrics.

Lin Zhang
School of Information Management
Wuhan University
Wuhan, China
zhanglin_1117@126.com

Lin Zhang is a Professor at the Wuhan University and a Guest Professor at both the Center for R&D Monitoring (ECOOM) at KU Leuven (since 2013) and the Chengdu Library and Information Center, CAS (since 2016). She obtained a PhD in Management Science from Dalian University of Technology (2010). Her research interests include subject clustering methods, interdisciplinarity measurement, and evaluation of individual scientists.

Xiaojun Hu
School of Basic Medical Sciences
Zhejiang University
Hangzhou, China
xjhu@zju.edu.cn

Xiaojun Hu is a Professor in Medical Informatics at Zhejiang University School of Medicine. She received a PhD in Public Administration from Zhejiang University. Her main interest lies in interdisciplinary research between scientometrics and biomedical informatics, including patent-metrics, and biological and disease networks. As a Principal Investigator, she has finished and undertaken research projects supported by the National Natural Science Foundation of China and Zhejiang Association for Science and Technology.

4. Google Scholar as a Data Source for Research Assessment

Emilio Delgado López-Cózar, Enrique Orduña-Malea, Alberto Martín-Martín

The launch of Google Scholar (GS) marked the beginning of a revolution in the scientific information market. This search engine, unlike traditional databases, automatically indexes information from the academic web. Its ease of use, together with its wide coverage and fast indexing speed, have made it the first tool most scientists currently turn to when they need to carry out a literature search. Additionally, the fact that its search results were accompanied from the beginning by citation counts, as well as the later development of secondary products that leverage this citation data (such as Google Scholar Metrics and Google Scholar Citations), made many scientists wonder about its potential as a source of data for bibliometric analyses. The goal of this chapter is to lay the foundations for the use of GS as a supplementary source (and in some disciplines, arguably the best alternative) for scientific evaluation. First, we present a general overview of how GS works. Second, we present empirical evidences about its main characteristics (size, coverage, and growth rate). Third, we carry out a systematic analysis of the main limitations this search engine presents as a tool for the evaluation of scientific performance. Lastly, we discuss the main differences between GS and other more traditional bibliographic databases in light of the correlations found between their citation data. We conclude that GS presents a broader view of the academic world because it has brought to light a great amount of sources that were not previously visible.

4.1 The Origins of Google Scholar

The development of the field of bibliometrics has always been reliant on the availability of large-scale sources of metadata about scientific publications, which are ultimately the raw materials used by bibliometricians to carry out their analyses [4.1]. The creation of the first citation indexes by Eugene Garfield (Science Citation Index in 1964, Social Science Citation Index in 1973, and the Arts and Humanities Citation Index in 1978) turned out to be a crucial turning point that enabled the development of modern bibliometric studies. His novel approach to bibliographic information systems opened the way to a completely new way of assessing scientific performance [4.2].

By indexing not only the articles published in scientific journals, but also the bibliographic references included in these articles, it was possible for the first time to track the relationships between scientists, journals, and institutions through the main tangible outputs these entities produce: scientific documents. Thus, the

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-02511-3_4) contains supplementary material, which is available to authorized users.

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_4

databases of the Institute for Scientific Information (now part of Clarivate's Web of Science; WoS) became the first, and for a long time the only available sources of data for bibliometric analyses, exercising an almost absolute monopoly in this field. The use of other specialized databases (such as MEDLINE, Chemical Abstracts, Inspec, or Biosis) for bibliometric purposes was testimonial, since they did not offer citation data nor other fields (e. g., full affiliations of the authors), which are vital to produce useful bibliometric reports.

It was not until the first decade of the twenty-first century that this monopoly was seriously challenged. Elsevier launched its Scopus citation database on November 3rd, 2004. Just a few weeks later (November 18th) GS was also launched. Scopus was conceived as a traditional subscription-based bibliographic database (which indexed a specific set of journals and conference proceedings) and was clearly a direct competitor of WoS. GS departed entirely from this approach, following instead the path of its big brother, the Google search engine, a decision that greatly impacted its design and coverage.

Simply put, GS is a specialized search engine that only indexes academic documents [4.3, 4]. GS's spiders constantly crawl the websites of universities, scientific publishers, topic and institutional repositories, databases, aggregators, library catalogues, and any other web spaces where they might find academic-like materials, regardless of their subject or language. GS indexes documents from the whole range of academic document types (books, book chapters, journal and conference articles, teaching materials, theses, posters, presentations, reports, patents, etc.). Unlike the cumulative and selective nature of WoS and Scopus, GS is dynamic; it reflects the state of the web as it is visible to its search robots and to the majority of users at a specific moment in time. Documents that for any reason become unavailable on the web will eventually disappear from GS too, as will the citations they provided to other documents [4.5].

GS, like the Google search engine before it, achieved instant success among users worldwide. The reason is easy to understand: GS finds most of the scientific information that circulates around the web in an easy and fast manner. Perhaps most importantly, it is free, unlike most of the bibliographic databases that existed before it, which are often only accessible through costly national or university-level subscriptions.

GS is currently the tool most users first turn to when they need to carry out a literature search. This has been evidenced by numerous studies [4.6–11]. *Bosman* and *Kramer*'s study [4.7] is the most recent and large-scale study on the matter. They conducted a survey on the changing landscape of scholarly communication between May 2015 and February 2016, obtaining more than 20 000 responses from researchers, university students, librarians, and other members of the scholarly community. With respect to the question *What tools do you use to search literature?* GS emerged as the preferred option, selected by 89% of the respondents, followed at a great distance by WoS (41%), Pubmed (40%), others (36%), and Scopus (26%).

Since its launch in 2004, GS's interface has gone through several renovations, but the really important changes (updates to its algorithms, coverage) usually happened under the hood, unbeknownst to most users. Some developments, however, did not go unnoticed. We are referring, of course, to the creation of its two secondary products: Google Scholar Citations (GSC) and Google Scholar Metrics (GSM). GSC was launched in July 2011 and provided a platform in which users could easily create an academic profile by pulling their publications from the data available in GS. Most interestingly, these profiles also displayed several author-level bibliometric indicators [4.12]. GSM was born on April 2012 as a ranking of scholarly publications according to their h-index calculated from GS data. This tool provides an easy way to identify the most influential publications (journals, proceedings, and repositories) and articles published in recent years [4.13].

Although these two tools never lose sight of GS's main purpose (they are intended to serve as search tools, one to find relevant researchers, the other to find influential articles and publications), they use bibliometric indicators as an evidence of relevance. For the first time, the GS team decided to put the citation data available in GS to other uses. Until the creation of those products, citation counts were only used as one of the parameters to rank documents in a search and a search aid for users (*Cited by* links in GS).

The availability of citation data in GS and its secondary products GSC and GSM has attracted the attention of some bibliometricians, and even scientists from other fields, who have realized that the data available in GS provides a much more comprehensive insight into the impact publications have on their respective academic communities than the data available in other citation databases. However, the use of GS for bibliometric purposes was never one of the applications GS's developers intended for this product, and so an exhaustive critical evaluation that analyzes its suitability for bibliometric analyses is necessary.

In order to do this, this chapter first presents a general overview of how GS and its secondary products GSC and GSM work, their inclusion policies, and how they respond (results offered) when specific stimuli (user queries) are applied. Second, we present empirical evidences regarding its size, evolution (growth rate, indexing speed), coverage (publishers, repositories, bibliographic databases, catalogues), and diversity (subjects, languages, document types). Third, we carry out a systematic analysis of the main limitations this search engine presents as a tool for the evaluation of scientific performance. Lastly, we discuss the main differences between GS and other traditional bibliographic databases in light of the correlations found between their citation data at the level of authors, documents, and journals.

The online version of this chapter is complemented by electronic supplementary material (ESM) in which the sources of the empirical analyses are made available, and further documents illustrating the findings on GS are provided [4.14].

4.2 Basic Functioning of Google Scholar

In this section, we first present a concise but accurate description of how the GS search engine works. Secondly, we describe its main inclusion criteria (both for sources and, especially, for documents). Lastly, we will briefly outline GSC and GSM.

4.2.1 The Academic Search Engine

Classic bibliographic databases usually work on the principle of whitelists. They first generate a whitelist of sources that meet some specific criteria (quality, subject scope, ...) and then index all the publications that appear in these sources. The historical tendency to select some specific sources (mainly journals) and not other channels for the dissemination of academic results (conference proceedings, books, reports, etc.) responds mainly for two reasons. First, it is a question of efficiency, usually referred to as Bradford's law of scattering [4.15], thanks to which we know that for any given topic, a small core of journals provides most of the articles on that topic. When faced with technological and economic constraints, maximizing returns by selecting only the core of journals that will be most useful for a given purpose seems a logical and natural response. The other reason has to do with the evaluative use of these databases. Due to their visibility and prestige, most authors want to publish their articles in these core journals, increasing the competition to get a manuscript accepted in these journals. The limited space for publication of the printing era, as well as the higher standards to which articles are held in these journals are what help project their image of prestige. To publish an article in a core journal is a difficult task, something that only the best researchers manage to do. In the same way, receiving a citation from an article published in a core journal also lends prestige to the cited article and its authors. This is the road to research excellence.

It goes without saying that this traditional approach (which prioritizes the optimization of resources and excellence) is not without its merits and has played an important role up until now. However, the irruption of GS represents a break from this paradigm. Unlike traditional bibliographic databases, which are selective by nature, GS parses the entire academic web, indexing every scholarly document it finds, regardless of its quality, and it does not differentiate between peer-reviewed and non-peer-reviewed content. GS is, rather, an academic search engine [4.3] with a bibliographic database that grows in a (mostly) unsupervised manner, and which has one clear purpose: facilitating the discovery of academic literature for everyone worldwide.

This unsupervised indexing process is possible thanks to the automated bots (sometimes also called spiders) that GS deploys throughout the web, similarly to how the Google search engine crawls the web as well. In GS, these bots are trained to locate academic resources, index their full texts (whenever possible), and extract their bibliographic descriptions (metadata). The process ends with the automated creation of a bibliographic record that is ready to be included in a search engine results page (SERP) when it is deemed relevant for a particular query. In order for a particular academic website to be successfully indexed in GS, certain technical requirements must be met, i. e., bots must be allowed to enter the website, there must be an easy-to-follow route to the article pages, and certain metadata must be available on these article pages. More detailed information can be found on the GS help pages [4.16].

When GS's bots are able to access the full text of the documents (either because the resource is openly available on the web, or thanks to the special agreements GS has with most commercial publishers), they also extract the list of cited references from each document. Thus, they are able to link citing and cited documents, which is how they can calculate citation counts.

When a list of cited references is parsed and processed, GS tries to find matches to those documents in its database. If it finds a match, it links the citing and cited document, and the cited document will have one more citation. However, if it does not find a match for a particular cited reference, the system will create a new bibliographic record of the type *[CITATION]*, to which the citing document will be linked. These records are also displayed in SERPs, although it is possible to exclude them, as their bibliographic information is often incomplete, and users will not be able to access their full text. A *[CITATION]* record can become a full-fledged record if GS finds another version of the document on the web (because someone deposits it on a repository, or it becomes available from a publisher, etc.), and merges the two versions. Needless to say, this entire process is also completely automated.

Lastly, GS considers a wide range of parameters for ranking documents in the SERPs, such as

> weighing the full text of each document, where it was published, who it was written by, as well as how often and how recently it has been cited in other scholarly literature. [4.17]

However, the detailed set of parameters and the weight each of them has in the ranking algorithm is not publicly available.

4.2.2 What Sources Does Google Scholar Index?

The previous section makes clear the distinction between GS and traditional bibliographic databases. In their own words, "we index papers, not journals" [4.18]. However, this statement is only partially true. We would rather define GS as a database that indexes web sources. Moreover, it deliberately includes some document collections (e. g., patents, court opinions, and *[CITATION]* type records).

GS crawls a wide variety of web sources and indexes everything from those sources that it identifies as academic documents. That is why GS includes all documents regardless of subject, language, country, or year of publication, and document type. The procedure GS follows to index new documents can be summarized in three steps, which are described below.

Step 1: Compilation of Sources

Over the years, GS has compiled a huge list of sources, ranging from websites of academic institutions (higher education institutions, national research councils, commercial publishers, private companies, professional societies, non-governmental organizations, etc.), to other discovery tools (bibliographic databases, catalogues, directories, repositories, other search engines) available across the web. These sources, which are the most likely to host academic content, are the ones that shape the academic web. Once they add a source to their private master list, GS's spiders will visit it periodically to check whether new documents have been added and also to verify that the documents indexed in the past are still available.

Besides the already mentioned academic sources, anyone can request that their website be considered for inclusion in GS. They are prepared to index websites that run in most of the common repository platforms (DSpace, EPrints, ...), journal platforms (OJS—open journal systems), and also simple personal websites.

Since GS's main objective is to facilitate content discovery, the sources must not require users to install additional applications, to log in, use Flash, JavaScript, form-based navigation, or any other kind of unreasonable methods to access the documents. In addition to that, the website should not display popups, interstitial ads, or disclaimers. They specifically state that

> all those websites that show log-in pages, error pages, or bare bibliographic data without abstracts will not be considered for inclusion and may be removed from GS.

Step 2: Document Types

The next step is to index the academic documents available in each source. GS does not index all the documents in a source, only those that are academic in nature.

GS states that they cover mostly "scholarly articles, journal papers, conference papers, technical reports, or their drafts, dissertations, pre-prints, post-prints, or abstracts". However, content such as "news or magazine articles, book reviews, and editorials" is not considered appropriate for GS. However, appropriateness is not really a constraint, and the documentation also states that "shorter articles, such as book reviews, news sections, editorials, announcements and letters, may or may not be included".

These ambiguous declarations are a consequence of the automated way in which the system operates. Let us see a practical example:

The University of Oxford's official website (http://ox.ac.uk) can be considered a reliable source of academic information. GS has added this web domain to its master list of indexable sources (like it does with most universities). However, not all documents hosted in ox.ac.uk are of an academic nature. GS needs to automatically differentiate academic documents from all the rest. To do this, the system applies two approaches:

1. The parser approach: GS uses full-text parsers to identify the structure of documents. Taking advantage of the fact that many academic documents tend to present a fairly standardized structure (title, author names, abstract, body of the article, references, ...), detecting whether or not a document is academic is often possible, although errors do occur.

2. The location approach: GS automatically indexes all the documents hosted in specific locations where it is reasonable to expect that all documents will have an academic nature, i. e., institutional repositories.

For this reason, despite what is stated in GS's documentation, it is possible to find a great range of document types in GS. Documents are usually stored in the HTML or PDF format. Parsing the structure from these documents is not enough to detect their specific typology (article, book chapter, conference paper, etc.) when additional metadata is not available. Moreover, once it has been decided that all content from a given source will be indexed, the actual document type stops mattering. For example, in GS we can find many book reviews, a document type that is explicitly considered inappropriate (Fig. 4.1).

Step 3: Documents

Lastly, documents themselves must also follow certain guidelines in order to be successfully indexed in GS. Some of them are compulsory, and others are only optional. Failure to comply with these rules may provoke an incorrect indexing of the documents, or, more likely, a complete exclusion from the search engine.

The system requires one URL per document (one intellectual work should not be divided into different files, while one URL should not contain various independent works). Additionally, the size of the files must not exceed 5 MB. Although documents of a larger size can appear in GS, their full text (including cited references) will be excluded if they do not comply with this rule.

HTML and PDF are the recommended file types. Other document types such as DOC, PS, or PPT are also indexed, but they are a very small minority and they might not be processed as effectively as the others. Additionally, PDFs must follow two important rules. First, all PDF files must have searchable text. If the PDFs are just scanned images, the full texts (including the cited references) will not be processed, since GS's crawlers are unable to parse images. Second, all URLs pointing to PDF files must end with the '.pdf' file extension.

There are also rules regarding the description of the articles through metadata. Some fields are compulsory for all documents (title, authors, and publication date), while others are specific to each document type. GS supports HTML meta tags in various formats: Highwire Press, Eprints, BE Press, PRISM, and Dublin Core (the last one as a last resort, since there are no specific fields for journal title, volume, issue, and page numbers in this format). If no metadata is readily available in the HTML meta tags of the page describing the article, GS will try to extract bibliographic information by parsing the full text of the document directly. For this reason, GS also makes recommendations regarding the layout of the full texts:

- The title, authors, and abstract should all be on the first page of the text file.

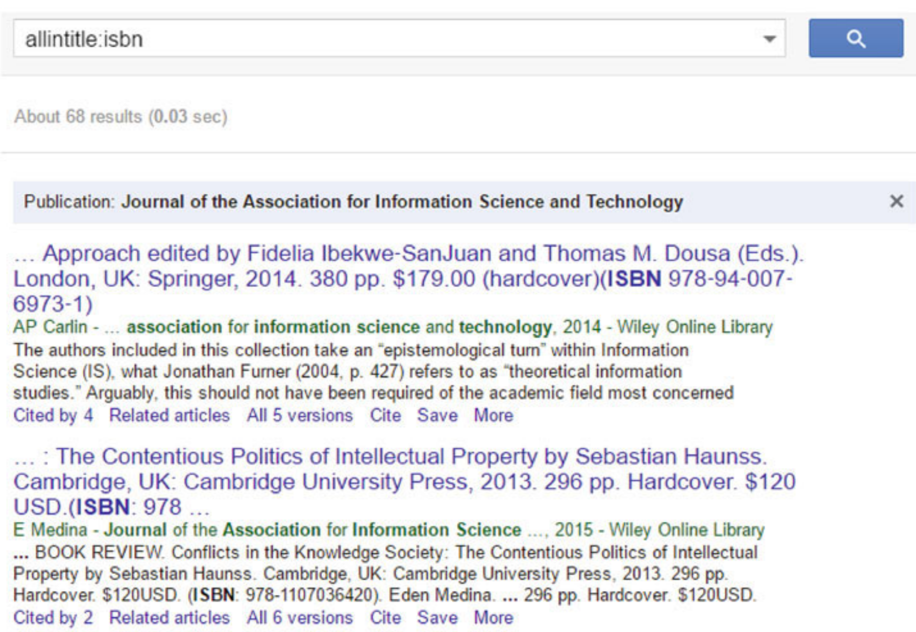


Fig. 4.1 Book reviews published in the *Journal of the Association for Information Science and Technology* (JASIST) and indexed in GS

- The title should be the first content in the document and no other text should be displayed with a larger font size.
- The list of authors should be listed below the title, in smaller font size, but larger than the font size used for the normal text.
- At the end of the document, there should be a separate section called *References* or *Bibliography*, containing a list of numbered references.

Lastly, the abstract of the document should be visible to all users that visit the article page (regardless of whether or not they have access to the full text of the document) without needing to click on any additional links or to log in. If this requisite is not met, it is likely that the document will not be indexed in GS.

4.2.3 Google Scholar's Official Bibliometric Products

GS has developed two secondary products that make use of both the bibliographic and citation data available in its core database. The first one (GSC) focuses on researchers, while the second one (GSM) focuses on journals and articles. This section describes each product briefly and discusses the bibliometric indicators they provide.

Google Scholar Citations

GSC was officially launched on November 16th, 2011. This tool is an academic profile service meant to help researchers maintain an up-to-date list of their publications without much effort (it is updated automatically), and it also facilitates searches of people (rather than documents) who are experts in any given academic topic.

First, GSC profiles contain structured personal information (name of the researcher, affiliation, and research interests). Second, the profiles show a list of all the publications written by the researcher. For each of the publications, both bibliographic (authors, title, source, year of publication) and citation data (number of citations and link to the list of citing documents) are offered. By default, documents are sorted decreasingly by number of citations, although they can also be sorted by year of publication or by title. Third, the profile also provides several author-level indicators (Table 4.1). These indicators are calculated considering two different time frames: first, without any time restriction (which is useful for comparisons of senior scholars), and second, considering only citations received in the last 5 years (which is useful for comparisons of early-career researchers). It is important to keep in mind that these indicators are calculated automatically from the data in the publication list, without any sort of human supervision.

In GSC, users have access to a document search tool that enables them to find their publications by means of author name searches (it is possible to search as many name variants as necessary), or by known document searches (usually title searches). After all documents have been found, researches may merge versions of the same document that GS has not been able to detect and fix bibliographic errors manually. All these operations only affect the researcher's profile and not other co-authors' profiles or the results in the GS search.

The platform also offers additional services, such as personalized alerts, lists of co-authors, areas of interests, list of authors by institution, etc. Therefore, authors can track the impact of their papers and other researchers' papers according to the data available in GS and be instantly informed of new papers published by other authors. All these features make GSC a powerful and free research monitoring system.

This product may be viewed as a first step in the transition from an uncontrolled database to a better-structured system where authors, journals, institutions, and areas of interest go through human filters. Nevertheless, the platform still lacks some essential features, like the identification of document types. This information can be defined by the owner of the profile on the document edit page, but it is not visible to other users visiting the profile. Another important issue is that author affiliations are not available at the level of documents. Although the affiliation field of the profile can be modified, it is only possible to display one affiliation at a time. However, researchers may change affiliations, and it would not be fair to ascribe all documents by a researcher to only one institution, if some of them were published while that researcher was working at other

Table 4.1 GSC's author-level metrics

Metrics	All	Last 5 years
Citations	Number of cites an author has received	Number of cites an author has received in the last 5 complete calendar years
h-Index	The largest number h such that h publications have at least h citations	The largest number h such that h publications have at least h new citations in the last 5 years
i10-Index	Number of publications with at least 10 citations	Number of publications with at least 10 citations in the last 5 years

institutions. Limitations like these diminish the usefulness of the platform for bibliometric studies (for more limitations, please see ESM, Appendix VI).

Google Scholar Metrics

GSM was launched on April 1st, 2012 and can be defined as a hybrid between a bibliographic and bibliometric product that presents a ranking of journals according to bibliometric indicators calculated using citation data from recently published articles in those journals. If GS represented a paradigm shift in the market of bibliographic databases, GSM accomplished something similar with respect to journal rankings, especially when compared to products like journal citation reports (JCR), the SCImago journal and country rank (SJR), or journal metrics [4.19]. GSM is an original product for various reasons:

- *Inclusion policies*: GSM only covers journals that have published at least 100 articles in the last 5 years and which have received at least one citation for any of those articles.
- *Coverage*: Apart from journals, GSM also covers some conference proceedings from computer science and electrical engineering, and preprint repositories. Other typologies like court opinions, books, and dissertations are explicitly excluded.
- *Sorting criteria*: Sources are sorted by their h5-index (h-index for articles published in a given 5-year period). The use of an h-index variant instead of a formula similar to the journal impact factor (JCR), SJR, SNIP (source normalized impact per paper), or CiteScore (journal metrics) is probably one of the most distinct features of this product. A description of the indicators available in GSM is presented in Table 4.2. For each journal, only the articles that contribute to the h5-index are displayed (h5-core). These articles are also accompanied by their citation counts, and the list of citing documents to each article is also available.
- *Categorization of sources*: The first variable of categorization is the language of publication. In the version available at the time of writing (launched in summer 2016, covering the period 2011–2015), the following languages were covered: English, Chinese, Portuguese, Spanish, German, Russian, French, Japanese, Korean, Polish, Ukrainian, and Indonesian. For each of these languages, except for English, the ranking displays the top 100 sources according to their h5-index. For English sources, a subject classification is also provided (Table 4.3).

Table 4.2 Google Scholar Metrics

Metrics	All
h5-Index	The largest number h such that at least h articles in that publication were cited at least h times each in the last 5 years
h5-Core	The set of articles from a journal with a citation count above the h5-index threshold
h5-Median	Median of the distribution of citations to the articles in the h5-core.

Table 4.3 Categories and number of subcategories in GSM

Categories	Number of subcategories
Business, economics, and management	16
Chemical and materials science	18
Engineering and computer science	58
Health and medical science	69
Humanities, literature, and arts	26
Life sciences and earth sciences	39
Physics and mathematics	24
Social sciences	52

  The classification scheme is made of 8 main categories and 302 subcategories. In each of the categories and subcategories, the number of results displayed is limited to the top 20 journals according to their h5-index.
  It should be pointed out that some subcategories are included in several categories (library and information science, for example, is included both in engineering and computer science, and in social sciences), and that one source may be included in more than one subcategory.
- *Search tool*: The platform also provides an internal search box, which enables users to locate journals that are not included in any of the general rankings. Users can carry out keyword queries, which will return sources with names that match the query. Each response to a query contains a maximum of 20 sources, also sorted according to their h5-index.

The peculiar features of this journal ranking have been tested, and numerous limitations found, such as a lack of name standardization, irreproducible data, and a questionable mix of publication typologies [4.20]. Nevertheless, the product has improved since its first editions, revealing itself as a potential source for the evaluation of journals in the areas of humanities and social sciences [4.21, 22].

4.3 Radiographing a *Big Data* Bibliographic Source

The goal of this section is to provide empirical data about the bibliographic properties of GS as a database. Three aspects will be discussed: size, coverage, and growth rate.

4.3.1 Size

One of the most crucial aspects that make us consider GS a *big data* source is the issue of determining its size. Unlike Scopus or WoS, highly controlled databases where finding out the total number of records only requires a simple query, GS is a search engine that presents what is available in the academic web at a specific moment in time. However, the web is not only dynamic, but also unstable and uncontrollable.

Therefore, the methodological difficulties of ascertaining the size of search engines are related to stability problems [4.23, 24], precision of the results obtained [4.25, 26], and the degree of permanence and persistence of the resources [4.27–29]. The high unstability of search results and the lack of precise search commands have led experts interested in finding the size of a search engine to use methods based on the extrapolation of frequencies of documents available in external sources [4.30].

As regards GS, there are two types of methods to calculate its size: direct methods (based on the execution of queries in the search engine) and indirect methods (estimations based on comparisons with external sources for which more information is known). Among the direct methods, three strategies are worth mentioning:

1. Web domain queries [4.31–33]
2. Year queries [4.3]
3. The so-called *absurd* queries [4.34].

Among the indirect methods, the capture/recapture method [4.35] and the proportion of documents in English respect to the total [4.34] have been attempted. For these last two methods, additional information must be known about the databases used as a reference. To date, these studies find that direct methods based on web domain queries and absurd queries are the ones that yield higher figures, which are similar in both cases.

Before discussing the calculation of the size of GS, it is appropriate to describe the characteristics of its coverage, since this is the key aspect to understanding the results. GS is currently made of two separate document collections: articles and case laws. The analysis of the latter is outside the scope of this chapter. The article collection is, in turn, divided into source documents and cited references (documents that GS's crawlers have only been able to find as references inside other source documents or certain metadata-only bibliographic databases). Cited references are marked with the text *[CITATION]* in SERPs and can also accrue citations of their own, which are displayed in the same way as for source documents. There is a last document type to which GS gives special attention in its interface: patents.

The integration of source documents and cited references in the same list of results breaks from the way WoS and Scopus handle these types of records, where each collection is displayed separately. In WoS, cited references are accessible from a completely separate search system (cited reference search), while in Scopus, cited references are also displayed separately as *secondary documents*.

There are two types of *[CITATION]* records: *Linked citations* (documents for which only basic bibliographic information—but no access to full-text—has been found in some library catalogue or metadata-only database) and *unlinked citations* (documents that have been cited in source documents and which the system has not been able to find anywhere else on the web).

Table 4.4 shows a compilation of the studies that have provided estimations of the size of GS. As can

Table 4.4 Compilation of studies on the size of GS

Authors	Date	Method	Coverage	Language	Size
Aguillo [4.32]	August 2010	Direct-domains	Articles + citations + patents	All	86 million
Ortega [4.3]	December 2012	Direct-date query	Articles + citations + patents	All	95 million
Khabsa and *Giles* [4.35]	January 2013	Indirect-cap/recap	N/A	English	99 million
Orduna-Malea et al. [4.34]	May 2014	Indirect-empirical studies	N/A	All	171 million
		Direct-date query	Articles + citations + patents	All	100 million
		Direct-absurd query	Articles + citations + patents	All	170 million
Aguillo [4.33]	January 2017	Direct-domains	Articles + citations + patents	All	194 million
Orduna-Malea et al. [4.36]	March 2017	Direct-absurd query	Articles + citations + patents	All	331 million
		–	Articles	All	184 million
		Direct-domains	Articles	All	197 million

N/A: not available

be expected, the results are affected by the estimation method, date of data collection, languages covered, and specific parameters of the searchers (inclusion or exclusion of cited references and patents).

Aiming to offer results as updated as possible, we replicated the absurd query in March 2017. A series of year queries, combined with the command 'fsdfsdgsdh.info' were carried out, and the number of hits each search yielded was collected. Through this simple procedure, we obtained a total of 184 001 450 source documents. Together with cited references (134 160 570) and patents (13 742 920), bringing the total to 330 804 940 documents.

Figure 4.2 offers a comparison of the size of GS, WoS core collection, and Scopus at two moments in time: 2011 and 2016. According to the most recent data, the coverage of GS seems to be almost three times as large as that of WoScc (2.8 : 1) and Scopus (2.7 : 1).

In order to test the robustness of the results, we compared the results returned by various types of queries to GS, carried out at different moments in time (Table 4.5). A high correlation was found among the results found for each year for all queries, regardless of whether or not cited references and patents were excluded.

Obviously, all these results are merely approximations of the size of GS. Its exact size cannot be ascertained with precision. Given the magnitude of the numbers the system returns (millions of documents), estimations are the best that we can expect when working with academic search engines.

Although the size of GS is clearly larger than that of other databases, the coverage by years can show us which database has a higher coverage in specific years. Figure 4.3 shows the number of documents by publication year that GS, WoS, and Scopus covered at the time of writing this.

As Fig. 4.3 shows, GS covers many more source documents than the other two databases, both for old material (first half of the nineteenth century) and recent

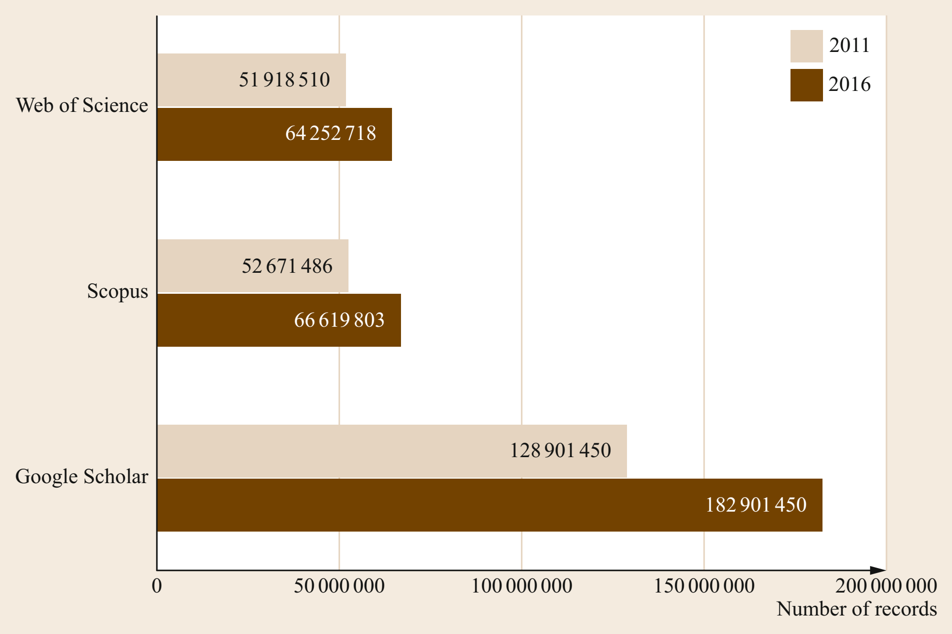


Fig. 4.2 Sectional coverage of GS, WoScc, and Scopus (from origin to 2016, included)

Table 4.5 Correlation among different queries (1800–2013)

Queries	Absurd pure (2017)	Date pure (2017)	Absurd full (2017)	Absurd full (2014)	Date full (2014)
Absurd-pure (2017)	1	0.997	0.992	0.978	0.976
Date-pure (2017)	0.997	1	0.994	0.984	0.983
Absurd-full (2017)	0.992	0.994	1	0.990	0.990
Absurd-full (2014)	0.978	0.984	0.990	1	0.995
Date-full (2014)	0.976	0.983	0.990	0.995	1

Pure: Excluding citations and patents
Full: Including citations and patents

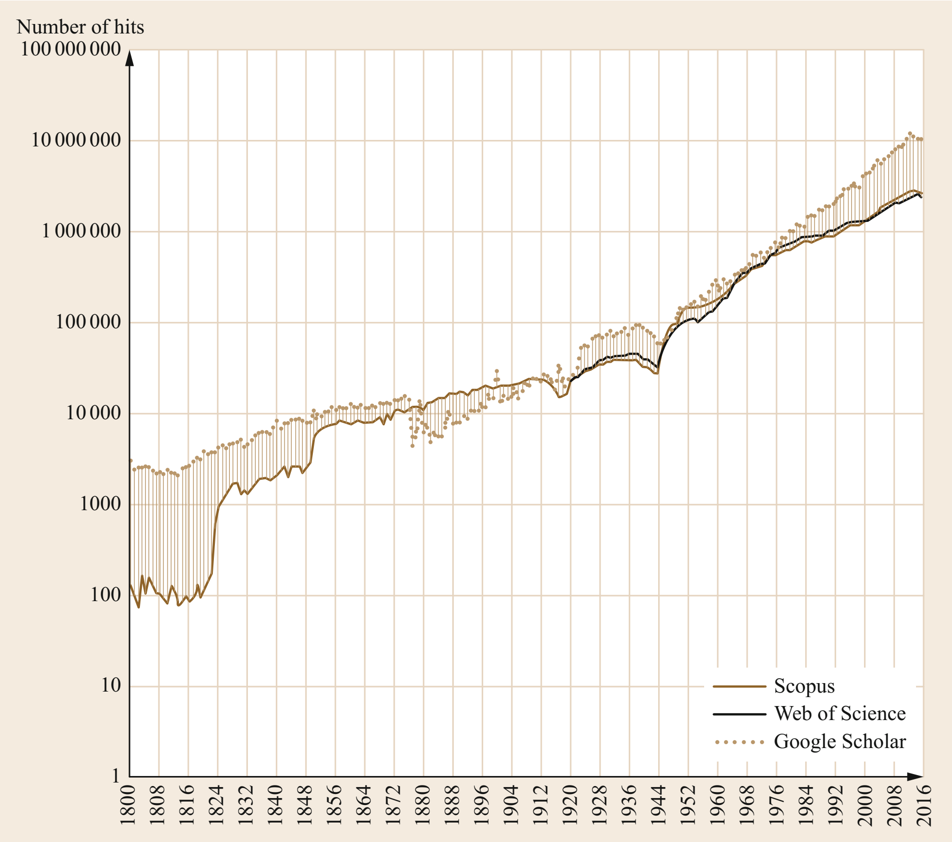


Fig. 4.3 Size of GS, WoScc, and Scopus (1800–2016). Note: The size of GS with respect to citing sources, eliminating cited references, in order to make GS comparable with the remaining databases

material (from the beginning of the twenty-first century onwards). The vast majority of the content covered by these databases was published in the current century (70.4% of all documents).

Up to this point, we have analyzed the size of GS by studying its source documents. However, it is also possible to approach this issue by studying the relationships between documents, that is, the citations themselves. As was previously said, an indirect method to find out the size of GS is to use other databases as a benchmark. Differences in citation counts in documents that are covered both by GS and the database used as a reference can be considered an indication of the underlying differences of their document bases: If a database has been able to find more citations than another database for a particular set of documents, its document base must be bigger (assuming the citation tracking mechanisms of the databases under comparison have roughly the same effectiveness).

Numerous studies have been published on this issue, most of them comparing GS to WoS and/or Scopus. The units of measurement used in these studies are varied:

1. Number of publications
2. Number of citations
3. h-indexes
4. Percentage of unique citations in each database.

A simple summary of these empirical results can reveal to which degree the size of GS's document base is bigger than that of the traditional databases, and how these differences have changed over time.

A table with a list of 63 empirical studies that address the issue of the size of GS as compared to WoS and/or Scopus is available in the electronic supplementary material [4.14, see Appendixes I, III, and IV]. For each study, the sample, discipline studied, unit of measurement used, results obtained, and ratio between the results for GS and the other databases are given. The results must be interpreted with caution, because the ratios depend to a large degree on the disciplines studied, the geographic scope (international or national), and the linguistic scope. Moreover, they can also be affected by the size of the samples (which, in general, tended to be very small).

From this meta-analysis, we can confirm several facts:

- To date, there are few studies that compare GS, WoS, and Scopus at the same time. The most frequent comparison is between GS and WoS.
- The vast majority of studies make comparisons on the basis of citations (54), way ahead of those that use documents—usually articles—(24), or indicators like the h-index (12).
- Out of the 63 studies, only 8 yielded results where WoS and/or Scopus surpassed GS in terms of size. All these studies analyzed STEM (science, technology, engineering and mathematics) fields, like chemistry. The majority of studies show that, for any given set of documents covered both by GS and WoS/Scopus, the first is able to find a higher number of citations.
- The greatest differences in favor of GS are found in the humanities and the social sciences. As regards STEM disciplines, GS is still able to find more citations, but the differences are less marked.
- The differences are greater when comparisons are made on the basis documents written in languages other than English.
- The differences between GS and the other databases seem to increase in the more recent studies, which might indicate an even broader coverage in GS respect to the other databases in recent years.
- Ratios of GS indicators to indicators from other databases are lower when comparisons are made on the basis of the h-index or the number of publications, and they are higher when citations are used instead. Ratios are even more in favor of GS when only unique citations are analyzed (citations found in one database and not in the others). These unique citations in GS, that is, its ability to find citations that no other database is able to find (not limiting itself to strictly scientific sources, but covering all the academic and professional sources it can find), are what make GS truly unique.

So as to provide a broader and more updated perspective on this issue, we have analyzed two large samples of documents covered both by GS and WoS. The results are consistent with the previous studies.

According to the samples analyzed in Table 4.6, the ratio of GS citations to WoS citations ranges from 1.54 (for a sample of 2.32 million articles and reviews published in 2009 or 2014) to 1.80 (for a sample of 69 261 highly-cited documents).

4.3.2 Coverage

After studying the size of GS, this section will focus on the characterisation of its content. To this end, its source, geographic, linguistic, discipline, and document type coverage will be discussed.

Source Coverage

Unfortunately, neither GS nor its secondary products (GSM and GSC) provide a master list of sources. As previously said, GS is not a journal database, but a service that indexes academic documents from many web domains. For this reason, efforts to determine its sources should try to identify these web domains first.

The first exhaustive study of the sources covered by GS based on web domains was carried out by *Aguillo* [4.32], who concluded that the most frequent geographic country code top-level domain (ccTLD) was '.cn' (China), and that Harvard University was the higher education institution that contributed more content to GS. *Ortega* [4.3] went a step further by estimating the proportion of content provided by several types of content providers:

1. Publishers (41.6%)
2. Other Google products (22.5%)
3. Subject repositories (16.9%)
4. Institutional repositories (11.8%).

Martín-Martín et al. [4.37] analyzed the sources of the primary versions of a set of 64 000 highly-cited documents in GS, finding close to 6000 content providers, among which the US National Institutes of Health (nih.gov), ResearchGate (http://researchgate.net), and Harvard University (harvard.edu) were the main providers of highly-cited documents. This study also found that generic top-level domains (TLDs) like

Table 4.6 Compilation of self-elaborated materials on GS/WoS citation ratios at the document level

Source	Date of data collection	Description	*N*	GS citations	WoS citations	Ratio GS/WoS
Self-elaborated. Publication forthcoming	June–October 2016	Articles and reviews with a DOI covered by WoS, published in 2009 or 2014. WoS data extracted from web interface	2.32 million	42.6 million	27.6 million	1.54
Self-elaborated for this chapter	February 2017	Highly-cited documents in master sample. WoS data extracted only from GS/WoS integration	69 261	80.8 million	44.9 million	1.80

'.edu', '.org', and '.com' were more frequent that geographic TLDs. Lastly, *Jamali* and *Nabavi* [4.38], based on a series of topic queries, used GS to estimate the sources (http://researchgate.net, http://nih.gov) and top-level domains ('.edu' and '.org') with a higher proportion of open-access documents.

Aiming to offer more updated results, we carried out a series of *site* queries in GS to find out the number hits returned for each of a list of 268 TLDs (251 geographic domains and 17 first-generation generic domains). The searches were carried out in March 2017. Publication year restrictions were not used, and cited references and patents were excluded.

Table 4.7 shows the main providers of documents according to their TLD. China is first among the geographic TLDs (12.12% of the total content), and commercial companies (.com) lead the list of generic domains (45.39% of the total content).

The sum of the results obtained for these 268 domains comes to 197 194 092 source documents, which is similar both to the one we obtained with the absurd query method in the previous section (184 001 450), and the one obtained by *Aguillo* [4.33], who used the same methodology (193 824 176). This reinforces the notion that the real number of source documents (excluding cited references) lies at around 200 million records. This figure is, however, a gross estimate, because there may be many duplicates, which will provoke an overestimation. At the same time, the fact that site command only counts primary versions would cause an infra-estimation if the web domain of the primary version does not match the web domain queried with the site search command.

Otherwise, the results indicate that most of the content is hosted on websites with generic (not geographic) TLDs (69.7%), undoubtedly because of the weight of journal publishers, standalone journals, and American universities (.edu).

For the purpose of delving deeper into the issue of source typologies, we proceeded to calculate the size of five types of web domains. We wanted to illustrate the diversity and weight of the different types of sources from which GS feeds: digital libraries and bibliographic information systems (Table 4.8), publishers (Table 4.9), learned and professional societies (Table 4.10), US government agencies and international organizations (Table 4.11), and universities (Table 4.12). These tables present the number of results found for each element, both including and excluding cited references. The goal of these tables is to enable us to observe which sources generate a higher quantity of bibliographic records in GS. Lastly, it is worth remembering that these results only consider the primary versions of the documents (those GS has selected as primary versions), and, therefore, these tables should not be understood as a ranking of sources sorted by size, but rather, a list that shows the diversity of sources available in GS.

The results in Tables 4.8–4.12 illustrate the main sources from which GS feeds: big bibliographic information systems, including databases (Pubmed, Europe

Table 4.7 Top 20 domain sources of GS (2017)

Rank	TLD	%	HCE	Description	Type
1	.com	45.39	89 500 000	Commercial	Generic
2	.org	16.38	32 300 000	Noncommercial	Generic
3	.cn	12.12	23 900 000	China	Country
4	.edu	3.55	7 010 000	US accredited postsecondary institutions	Generic
5	.jp	3.40	6 700 000	Japan	Country
6	.net	2.06	4 070 000	Network services	Generic
7	.ru	1.72	3 400 000	Russian Federation	Country
8	.gov	1.69	3 340 000	US Government	Generic
9	.br	1.35	2 670 000	Brazil	Country
10	.fr	1.22	2 400 000	France	Country
11	.kr	0.94	1 850 000	Korea Republic of	Country
12	.ua	0.69	1 360 000	Ukraine	Country
13	.id	0.65	1 280 000	Indonesia	Country
14	.es	0.63	1 250 000	Spain	Country
15	.pl	0.56	1 110 000	Poland	Country
16	.de	0.51	1 010 000	Germany	Country
17	.au	0.44	864 000	Australia	Country
18	.uk	0.44	863 000	United Kingdom	Country
19	.it	0.40	797 000	Italy	Country
20	.ca	0.37	734 000	Canada	Country

Notes: Cited references excluded, HCE: Hit count estimate

Table 4.8 Digital libraries (DL) and bibliographic information systems (BIS) sources of GS

Rank	DL and BIS	URL	Source Type	Hits	
				Citations excluded	Citations included
1	China National Knowledge Infrastructure	http://cnki.com.cn	Database	17 600 000	19 600 000
2	Google books	http://books.google.com	Engine search	3 860 000	9 300 000
3	JSTOR	http://jstor.org	Digital library	2 920 000	4 680 000
4	Europe PubMed Central	http://europepmc.org	Subject repository	2 290 000	4 310 000
5	ResearchGate	http://researchgate.net	Social network	2 020 000	2 040 000
6	Proquest	http://proquest.com	Database	1 670 000	1 750 000
7	Astrophysics data system	http://adsabs.harvard.edu	Database	1 510 000	2 040 000
8	J-STAGE	http://jstage.jst.go.jp	E-journal aggregator	1 460 000	1 750 000
9	Pubmed	http://ncbi.nlm.nih.gov	Subject Repository	1 360 000	3 350 000
10	Cyberleninka	http://cyberleninka.ru	Digital library	1 150 000	1 200 000
11	CAB direct	http://cabdirect.org	Database	1 100 000	1 100 000
12	Refdoc	http://cat.inist.fr	Database	1 080 000	2 390 000
13	Academia.edu	http://academia.edu	Social network	1 020 000	1 030 000
14	CiteSeerX	http://citeseerx.ist.psu.edu	Search engine	1 010 000	997 000
15	ERIC	http://eric.ed.gov	Database	635 000	695 000
16	AGRIS	http://agris.fao.org	Database	537 000	3 620 000
17	Semantic scholar	http://semanticscholar.org	Search engine	526 000	527 000
18	EBSCO	http://ebscohost.com	Database	479 000	479 000
19	Dialnet	http://dialnet.unirioja.es	Bibliographic portal	458 000	2 280 000
20	ARXIV	http://arxiv.org	Subject repository	403 000	407 000

[a]ERIC: Education Resources Information Center

Table 4.9 Publisher sources of GS

Rank	Publishers	URL	Hits	
			Citations excluded	Citations included
1	[a]Elsevier 2	http://sciencedirect.com	9 340 000	9 410 000
2	John Wiley and Sons	http://wiley.com	5 960 000	5 970 000
3	Springer	http://springer.com	5 590 000	5 770 000
4	Taylor and Francis	http://tandfonline.com	3 200 000	3 240 000
5	Sage	http://sagepub.com	1 370 000	1 560 000
6	Lippincott Williams and Wilkins	http://lww.com	1 240 000	1 240 000
7	Cambridge University Press	http://cambridge.org	1 100 000	1 270 000
8	[b]Oxford University Press 2	http://oxfordjournals.org	951 000	1 240 000
9	Walter de Gruyter	http://degruyter.com	595 000	622 000
10	Nature Publishing Group	http://nature.com	428 000	458 000
11	Karger Publishers	http://karger.com	347 000	348 000
12	Chemical Abstracts Service	http://pubs.acs.org	325 000	325 000
13	BioMed Central	http://biomedcentral.com	279 000	279 000
14	Emerald	http://emeraldinsight.com	236 000	259 000
15	PLoS	http://journals.plos.org	203 000	203 000
16	World Scientific Publishing	http://worldscientific.com	168 000	170 000
17	Hindawi	http://hindawi.com	167 000	193 000
18	Elsevier 1	http://elsevier.com	108 000	192 000
19	Inderscience Publishers	http://inderscienceonline.com	82 500	82 500
20	Brill	http://booksandjournals.brillonline.com	68 100	122 000

[a] This publisher owns another web domain (elsevier.com), in which we obtained 108 000 additional documents
[b] This publisher owns another web domain (oup.com), in which we obtained 4290 additional documents

Table 4.10 Learned and professional societies of GS

Rank	Learned and professional societies	URL	Hits	
			Citations excluded	Citations included
1	Institute of Electrical and Electronics Engineers (IEEE)	http://ieee.org	3 410 000	3 650 000
2	Institute of Physics (IOP)	http://iop.org	667 000	702 000
3	Royal Society of Chemistry (RSC)	http://rsc.org	470 000	476 000
4	Association for Computing Machinery (ACM)	http://acm.org	447 000	601 000
5	American Psychological Association (APA)	http://apa.org	406 000	448 000
6	American Chemical Society (ACS)	http://acs.org	326 000	327 000
7	American Society of Microbiology	http://asm.org	253 000	256 000
8	International Union of Crystallography	http://iucr.org	120 000	122 000
9	American Mathematical Society (AMS)	http://ams.org	93 500	112 000
10	American Meteorological Society (AMS)	http://ametsoc.org	59 100	66 300

Table 4.11 Government agencies and international organizations sources of GS

Rank	Government agencies and international organizations	URL	Hits	
			Citations excluded	Citations included
1	National Institute of Informatics (NII)	http://nii.ac.jp	2 960 000	13 300 000
2	Japan Science and Technology Agency (JST)	http://jst.go.jp	2 740 000	3 060 000
3	US National Institute of Health (NIH)	http://nih.gov	1 420 000	3 380 000
4	Institut de l'information scientifique et technique	http://inist.fr	1 100 000	2 420 000
5	US Department of Education	http://ed.gov	636 000	696 000
6	Office of Scientific and Technical Information	http://osti.gov	635 000	1 040 000
7	Food and Agricultural Organization (FAO)	http://fao.org	545 000	3 510 000
8	Defense Technical Information Center	http://dtic.mil	505 000	644 000
9	National Aeronautics and Space Administration (NASA)	http://nasa.gov	205 000	252 000
10	National Criminal Justice Reference Service	http://ncjrs.gov	120 000	124 000

Table 4.12 University sources of GS

Rank	Universities	URL	Hits	
			Citations excluded	Citations included
1	Harvard University	http://harvard.edu	1 410 000	2 170 000
2	Pennsylvania State University	http://psu.edu	1 030 000	1 080 000
3	Universidad de La Rioja	http://unirioja.es	442 000	2 280 000
4	University of Chicago	http://uchicago.edu	329 000	346 000
5	Johns Hopkins University	http://jhu.edu	324 000	340 000
6	Universidade de São Paulo USP	http://usp.br	155 000	197 000
7	Masarykova Univerzita v Brně	http://muni.cz	121 000	125 000
8	Universiteit van Amsterdam	http://uva.nl	105 000	108 000
9	Universidad Complutense de Madrid	http://ucm.es	105 000	356 000
10	Helsingin yliopisto	http://helsinki.fi	91 400	142 000

Pubmed Central, ADS), big commercial publishers (Elsevier, Springer and Wiley in particular), other academic search engines (Semantic Scholar, Citeseer, ...), subject repositories (arXiv.org), social platforms (ResearchGate, Academia.edu), as well as Google's own book platform (Google Books). It additionally includes, research government agencies (like the Japanese National Institute of Informatics and the Japan Science and Technology Agency), professional associations (IEEE), and universities (Harvard University is still leads this group). It should be noted that the results of some institutions (especially universities) are influenced by the existence of bibliographic products that are hosted within these universities' domains (Dialnet at the University of La Rioja, CiteseerX at Pennsylvania State University, AGRIS in FAO, ERIC in the US Department of Education, etc.).

However, these results should be interpreted with caution, because the methodology used to collect the data has several limitations: All hit counts displayed by GS (and Google, for that matter) are only approximations, not exact figures. What is more, the *site*: command is not exhaustive either; it only works with the primary versions of the documents in GS. GS implements a procedure to group together all the versions of a same document that may be available on the web [4.39]: subject and institutional repositories, the author's personal website, a social platform, and the official *version of record* available in a journal or on publisher's website. From all the versions found by the search engine, one of them (usually the publisher version, if there is one) is selected as the primary version. The rest of the versions can be found under the *All x versions* link available below each record.

This means that when a search containing the site: command is carried out, the system will only return the records in which the source of the primary version matches the searched source, even though there may be many more records from that source that have not been considered primary versions. If we focus on the case of ResearchGate, we can see that the 2 020 000 documents found (Table 4.8) are very far from the over 100 million documents the company claimed to cover in March 2017 [4.40]. This difference can be explained in part by the documents that are covered as secondary and not primary versions. For these reasons, the results in Tables 4.8–4.12 are most likely an underestimation of the real coverage, although it provides important clues as to the main sources it indexes.

Geographic Coverage

The geographic coverage of the documents covered by GS is also difficult to analyze, because the system is not designed to carry out searches based on authors' institutional affiliations. Therefore, like with the source coverage, a possible, although biased, approach is to analyze the distribution of geographic domains. *Orduna-Malea* and *Delgado López-Cózar*, using this methodology [4.41], found that the domains for the United States, China, and Japan were the ones that yielded the highest hit counts estimates (HCE), which is consistent with the results obtained by *Aguillo* [4.32]. Table 4.13 shows the top ten geographic domains according to their HCE, and their evolution in the last 6 years [4.32, 41].

This strategy, however, is imprecise, because it does not consider generic TLDs like '.com' and '.org', which are precisely the ones that are most used. This explains why the United States are clearly under-represented, since a great proportion of '.com' domains belong to institutions from this country [4.42].

There are few studies on the geographic distribution of documents in GS, and those few that have been published focus on the geographical origin of the journals indexed in GSM, and the comparison of these journals with the ones covered by WoS and Scopus in specific disciplines like communication [4.43], nursing [4.44], and library and information science [4.45] (Fig. 4.4). GS and GSM seem to get closer to the actual distribution of scientific journals by country of publication than the other databases.

Similar results were found in a study of over 9000 arts, humanities, and social sciences journals indexed in GSM, where GSM not only covers journals from more countries, but the English-language bias is clearly less pronounced than in Scopus and WoS (Fig. 4.5).

Linguistic Coverage

It is also difficult to find out the language distribution of the documents covered by GS, because the

Table 4.13 Top ten geographic domains of GS

Country	TLD	Hit counts estimate			(%)
		2010	2013	2016	
China	.cn	7 520 000	30 700 000	23 900 000	12.12
USA	–[a]	N/A	16 019 000	10 943 500	5.55
Japan	.jp	1 720 000	10 400 000	6 700 000	3.40
Russia	.ru	995 000	N/A	3 400 000	1.72
Brazil	.br	1 440 000	2 320 000	2 670 000	1.35
France	.fr	2 820 000	4 210 000	2 400 000	1.22
South Korea	.kr	481 000	1 720 000	1 850 000	0.94
Ukraine	.ua	210 000	N/A	1 360 000	0.69
Indonesia	.id	N/A	N/A	1 280 000	0.65
Spain	.es	907 000	2 990 000	1 250 000	0.63
Poland	.pl	220 000	N/A	1 110 000	0.56

2010 data: from [4.32]
2013 data: from [4.41]
2016 data: self-elaborated for this chapter. In all cases, citations are excluded; N/A: not available
[a] US data is obtained by merging the results from '.us', '.mil', '.edu', and '.gov' web domains

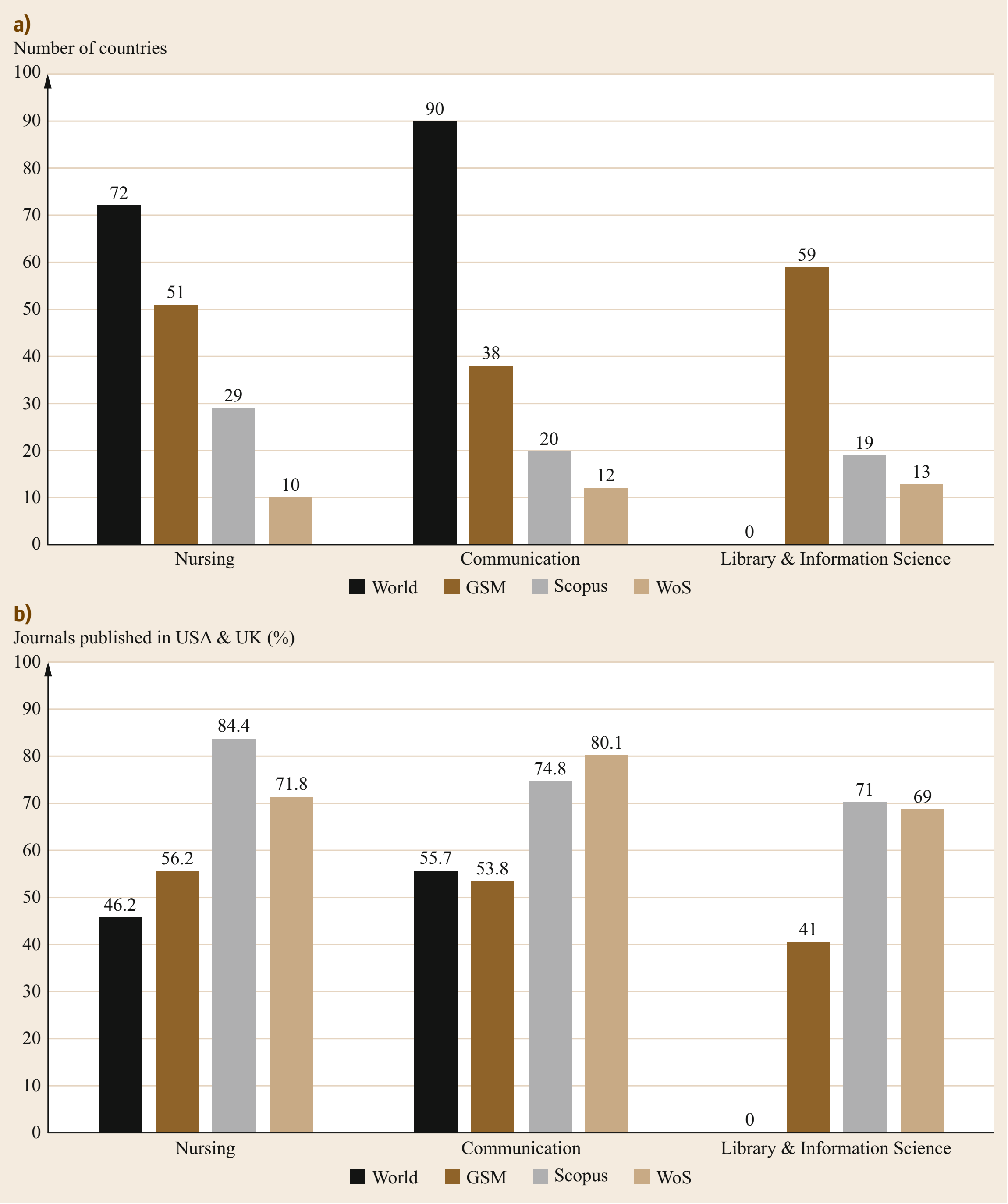


Fig. 4.4 Nursing, communication, and library and information science journals published according to GSM, Scopus, and WoS data **(a)**. Number of countries where the journals are published **(b)**. Percentage of journals that are published in the USA or UK. Note: In the case of library and information science journals, GS is used instead of GSM; and no data is available for the world category

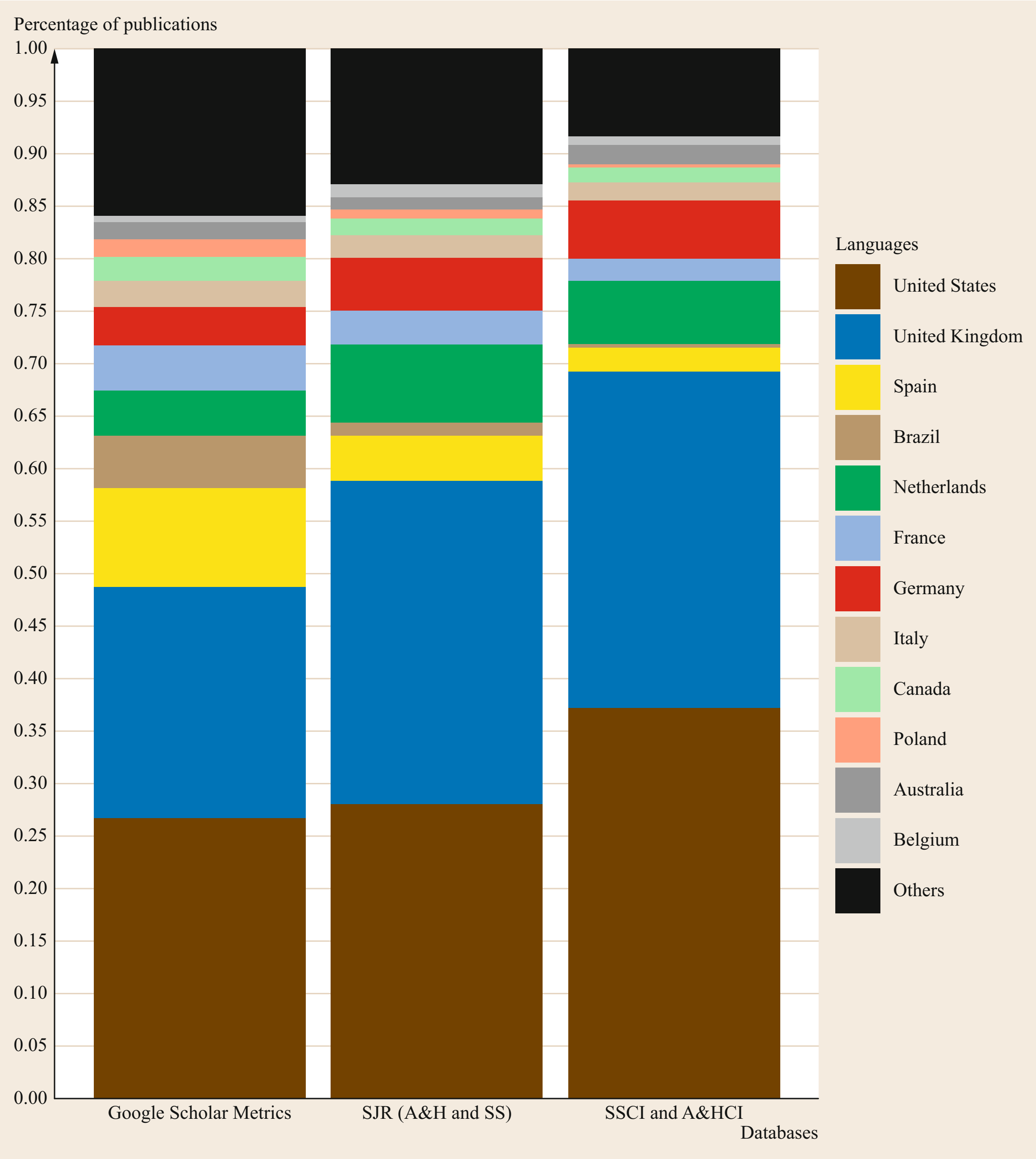


Fig. 4.5 Distribution of countries of publication in arts, humanities, and social sciences journals covered by GSM, SJR (Scopus), and JCR (WoS)

options the interface offers to filter by language are very limited. Users can limit search results to one or several of the following languages: Chinese (simplified and traditional), Dutch, English, French, German, Italian, Japanese, Korean, Polish, Portuguese, Spanish, and Turkish. However, users have to navigate to the settings page to find these options; they are not available from the main search interface.

In a study carried out by *Orduna-Malea* and *Delgado López-Cózar* [4.41], which analyzed the quantity of records by language in WoS, Scopus, and GS, a very high percentage of documents written in English was found in the first two databases (90%), while GS offered a higher linguistic diversity, because it covered other languages (especially Italian, Spanish, French, and Japanese). A year later, *Orduna-Malea* et al. [4.34], based on an analysis of the empirical studies on this topic, estimated that documents in the English language represented approximately 65% of all documents in GS.

Studies on the publication languages of journals available in GSM, as compared to those available in WoS or Scopus, reach the same conclusions. Both in communication journals [4.43], nursing journals [4.44], as well as in library and information science journals [4.45], GSM is closer to the actual distribution of languages used in scientific journals around the world, thus overcoming the bias towards English-language sources that prevails in Scopus and WoS (Table 4.14). While in the latter two, the proportion of English-language sources ranges from 80 to 93% of all sources, in GSM this figure is much lower: between 61 and 65% (Table 4.15).

An analysis of the arts, humanities, and social sciences journals available in the 2010–2014 edition of GSM yields similar results: GSM not only covers journals written in more languages, but its English-language bias is also lower than in the other two databases (Fig. 4.6), in spite of the fact that the study focused only on journals with titles written in Latin characters. GSM, because of its inclusion criteria, does not cover many journals for which articles are available in GS.

Aiming to obtain additional empirical data about the linguistic distribution of the content available in GS, we carried out a series of searches in GS. For each of the 12 languages that GS allows users to choose from to limit the search results (simplified Chinese, traditional Chinese, Dutch, English, French, German, Italian, Japanese, Korean, Polish, Portuguese, Spanish, and Turkish), 67 keyword-free, publication year queries were carried out, one for each publication year for the period 1950–2016 (871 queries in total).

Table 4.14 Number of different languages in nursing and communication journals indexed in GSM, Scopus, and WoS

Discipline	World	GSM	Scopus	WoS
Nursing	33	20	13	6
Communication	23	14	7	6

Table 4.15 Percentage of journals published in English in nursing and communication, and indexed in GSM, Scopus, and WoS

Discipline	World	GSM	Scopus	WoS
Nursing	57.2	61.9	81.2	92.7
Communication	70	65.3	91.6	87.8

Table 4.16 shows the distribution of results by language (excluding cited references and patents). As can be seen, English-language results make up half of the total amount of results (49.8%), followed by the sum of simplified and traditional Chinese results (33.7%). These results are consistent with the figures on geographic coverage through the analysis of web domains presented previously and confirm the preeminence of the United States and China in GS's coverage.

Even considering the disproportionately huge amount of English and Chinese results (which are clearly influenced by the sources from which GS extracts data), the distribution of the other languages is unquestionably more varied than the distribution presented by other databases. Figure 4.7 shows the relative distribution of languages in GS, Scopus, and WoS.

While the percentage of documents published in English in WoS and Scopus is of 90 and 80%, respectively, in GS the percentage is closer to 50%, and therefore the rest of languages are noticeably better represented.

Again, we would like to warn that the results obtained must be interpreted in the context of a search engine, and not a bibliographic database. GS automatically identifies the language of a document from certain parameters. However, a document might contain text in several languages. Therefore, the same document might be classified in various languages. Additionally, in some cases, the fact that a document is hosted in a geographic domain can help GS identify its language (for example, '.cn' is associated with the Chinese language), even if sometimes the documents are not written in the expected language [4.46]. For those reasons, the number of results might be an overestimation.

Table 4.16 Distribution of languages in GS results (cited references and patents excluded)

Language	Documents	(%)
English	90 932 140	49.76
Chinese	61 545 203	33.70
Japanese	6 327 073	3.46
German	4 326 244	2.37
Spanish	4 144 354	2.27
French	3 657 705	2.00
Portuguese	2 403 898	1.32
Korean	2 131 744	1.17
Italian	999 134	0.55
Polish	766 266	0.42
Dutch	475 703	0.26
Turkish	472 830	0.26
Other	4 534 156	2.48
Total	182 716 450	100

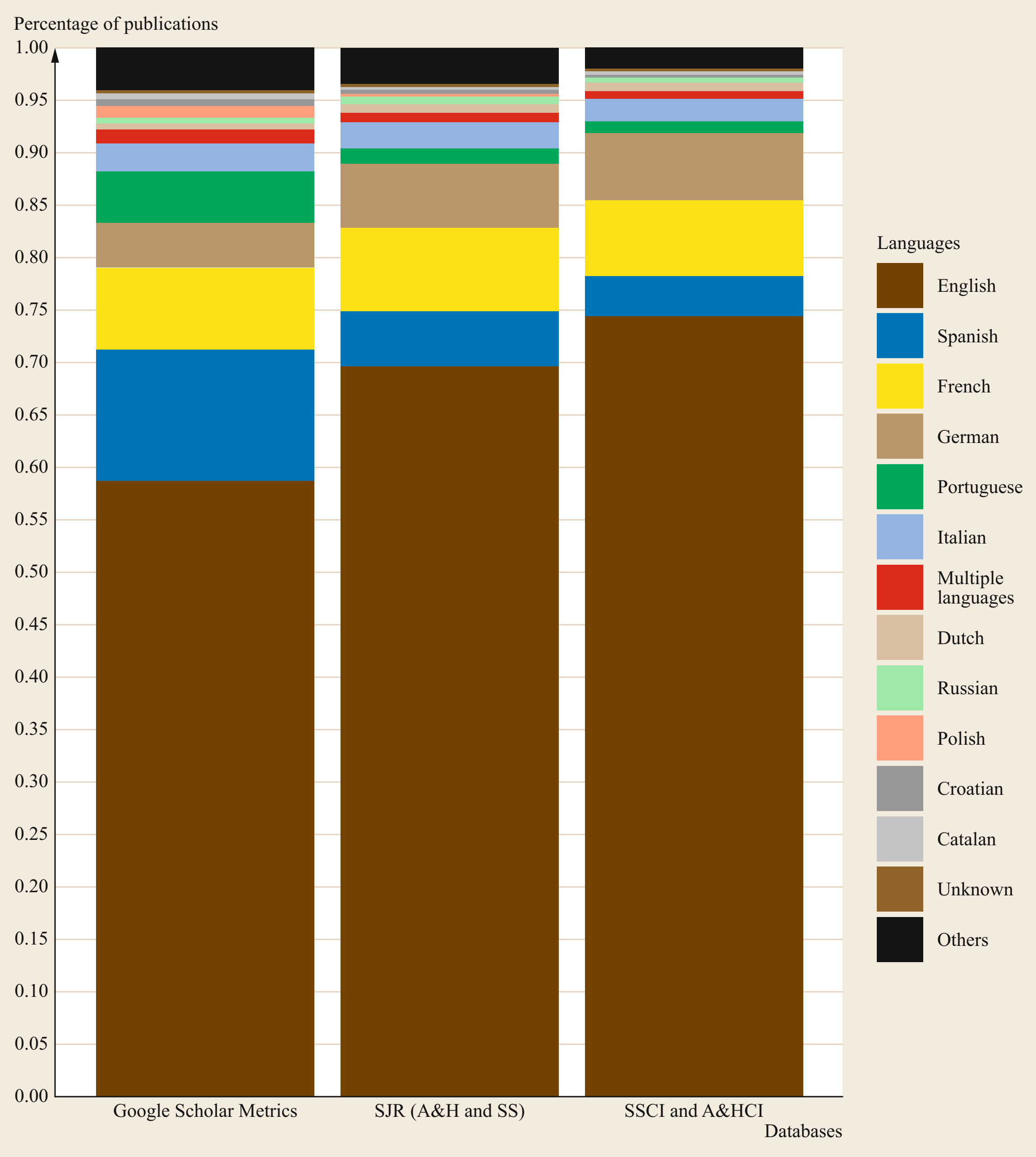


Fig. 4.6 Distribution of arts, humanities, and social sciences journals indexed in GSM (2010–2014) as compared to SCImago journal Rank (arts and humanities (A&H) and social sciences (SS) only), and journal citation reports (SSCI and A&HCI (Arts and Humanities Citation Index))

Discipline Coverage

Discipline coverage is another crucial aspect of the analysis of a bibliographic database. However, studies published to date mostly deal with journal coverage in GS as regards specific disciplines and countries.

From the data available about Spanish journals in the areas of social sciences, covered by GS and GSM in the year 2011 [4.47], and the data available in IN-RECS [4.48], a ranking of Spanish journals in disciplines related to the social sciences, a very informative table about the coverage of GS and GSM was developed (Table 4.17). Of the 1090 Spanish journals studied in IN-RECS, 95.2% (1038) were covered by GS. What is more, for some disciplines, GS covered even more

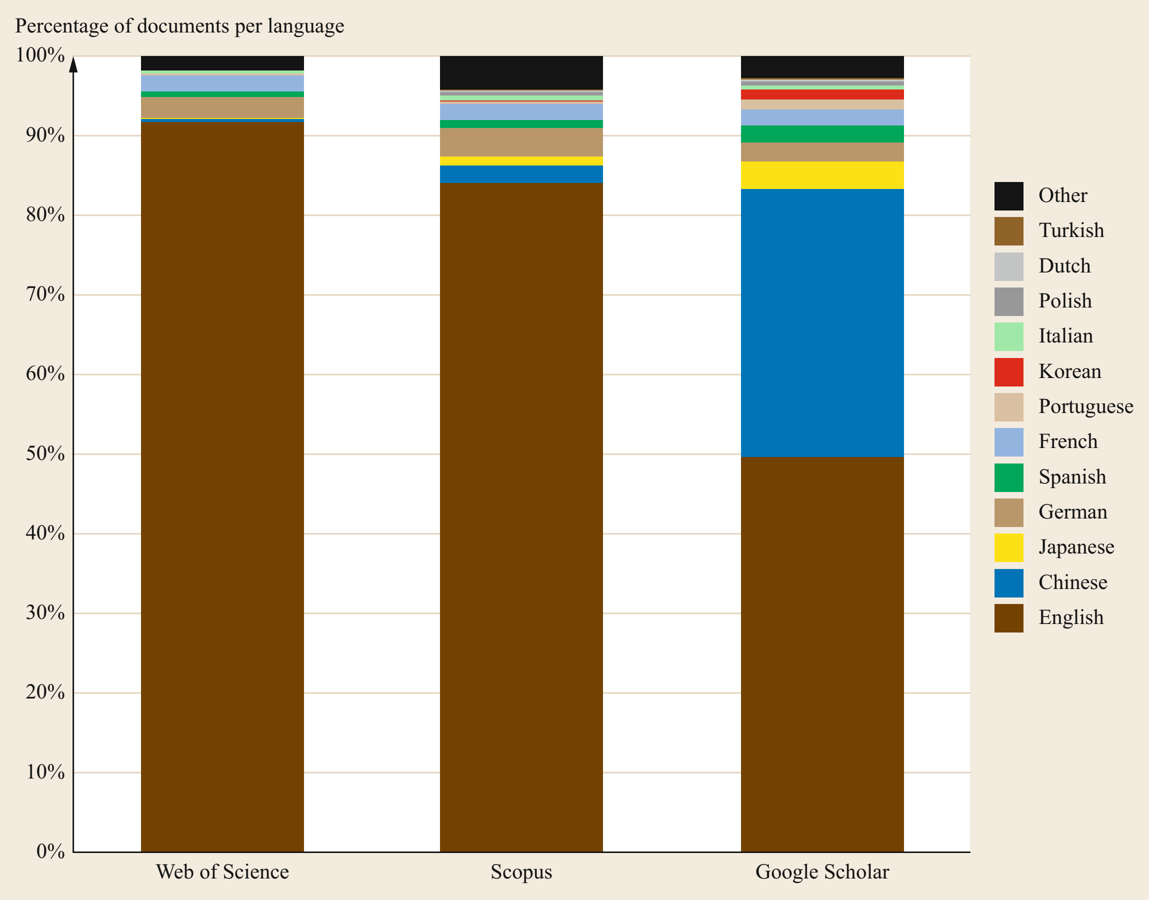


Fig. 4.7 Distribution of the languages of documents indexed in GS, Scopus, and WoS (1800–2016)

Table 4.17 Coverage of Spanish journals by discipline in IN-RECS, GS, and GSM

Discipline	IN-RECS	GS		GSM	
	Year (2010)	Year (2011)	Coverage (%)	Year (2011)	Coverage (%)
Law	341	251	74	110	43.8
Education	166	157	95	69	43.9
Economy	136	137	101	55	40.1
Psychology	108	109	101	42	38.5
Sociology	82	87	106	25	28.7
Political science and administration	60	56	93	21	37.5
Geography	51	54	106	15	27.8
Anthropology	46	46	100	10	21.7
Sport	N/A	42	N/A	14	33.3
Urban studies	43	39	91	15	38.5
Library and information science	33	36	109	15	41.7
Communication	24	24	100	12	50.0
Total	1090	1038	95	403	37.0

journals than IN-RECS. If we extrapolate those results, we can estimate that GS is very close to covering all active scientific journals. Of course, this hypothesis would require a more varied sample of journals to be tested.

On the other hand, GSM covers just over a third of all active Spanish social sciences journals (36.97%). This is most likely caused by the inclusion criteria enforced by the system (journals must have published

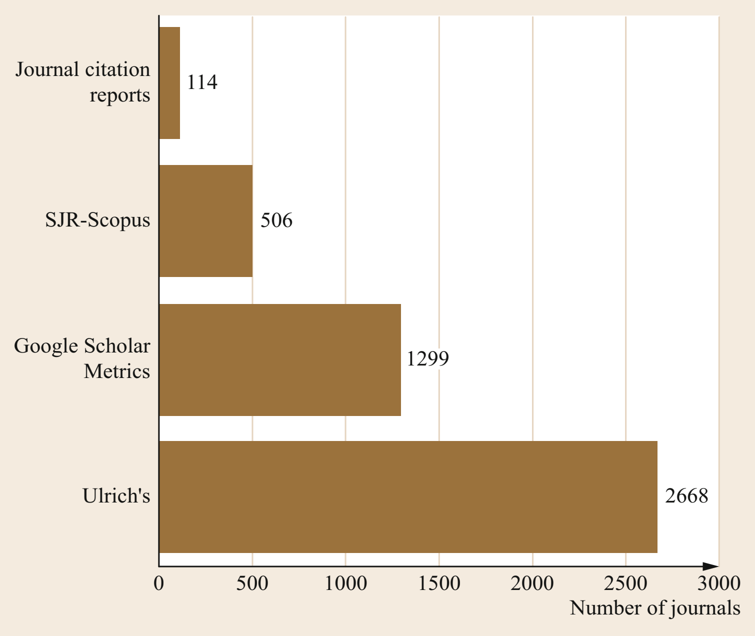


Fig. 4.8 Number of Spanish scientific journals covered by journal citation reports, Scopus, GSM, and Ulrich's directory

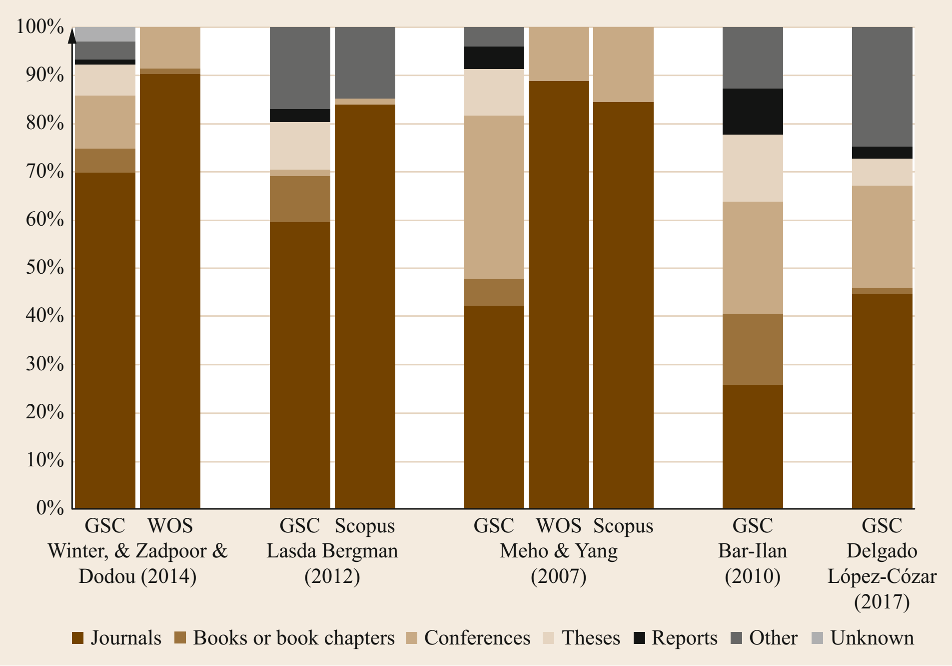


Fig. 4.9 Comparison of the document type distributions in GS, WoS, and Scopus as found in several studies

at least 100 articles in the last 5 years and have received at least one citation). Nevertheless, even with those limitations, GSM still covers many more journals than traditional databases (WoS and Scopus), and, therefore, is able to display a much broader spectrum of disciplines, better representing the scientific landscape.

In a study focused on the quantity of Spanish journals indexed in GSM, as compared with the total number of active Spanish journals, which according to the Ulrich's directory is around 2668 journals [4.49], similar results are obtained. Only 48.7% (1299) of Spanish journals were found in GSM (Fig. 4.8), but if this result is compared to the traditional journal rankings, we find that, in spite of its inclusion criteria, GSM covers twice the amount of journals than SJR (Scopus data), and ten times more than the journal citations reports (WoS data). However, there are studies that yield different results. *Gu* and *Blackmore* [4.50] find that, of a sample of 41 787 refereed academic journals from all disciplines covered by the Ulrich's directory, only 20.8% (10 354) were to be found in GSM, a lower amount than those that were found in SJR (32%, 15 911).

An additional method to quantify the number of publications covered by databases is to analyze the number of citations received by a sample of documents in a specific scientific discipline. With this approach, we can find studies focused on journals [4.51, 52] and researchers [4.53–56]. There is one other approach, based on the characterization of the documents returned by GS to sets of topic queries [4.38].

Important disciplinary differences can be observed in the list of studies that offer empirical evidences on the functioning of GS. While in the social sciences, the humanities, and engineering (especially in computer science) the ratio of citations in GS to citations in other sources is very high, the differences are much lower when STEM fields are analyzed. As was previously commented, the field of chemistry was initially covered poorly in GS, mostly due to the refusal of the American Chemical Society and other important publishers to be indexed by the search engine. These problems have already been solved [4.53, 57].

Document Type Coverage

Lastly, the last aspect of the coverage of GS has to do with document typologies. In order to learn about the wealth and diversity of document sources from which GS extracts data, we must summarize the results offered by the studies that have analyzed the distribution of citing documents according to their typology in several databases (Fig. 4.9). The main feature of GS is that it indexes more diverse document typologies. Indeed, it is the database where journal articles constitute a lower percentage of the total documents (from 28 to 70% depending on the samples). Conversely, in WoS and Scopus, journal articles make up 90% of the documents, which means there is a very limited coverage of conference communications and books.

Aiming to obtain more updated and representative results, a total of 871 queries (by language and year of publication) were carried out (including cited references and patents in the search), extracting the 1000 search results available for each query. There were 861 843 documents extracted [4.58].

Given that GS does not provide information about the typology of the documents that are displayed in

Table 4.18 Distribution of document typologies in GS ($N = 861\,839$ documents)

Document type	TOT	EN	CN	JP	DE	ES	FR	PT	KO	IT	PL	NL	TR
Unknown	463 290	11 299	85 547	29 313	32 092	40 926	27 375	38 655	26 206	38 773	41 803	44 671	46 630
Article	260 211	39 264	38 550	34 416	16 087	10 384	21 357	15 998	33 606	7584	13 638	13 499	15 828
Book/Chapter	120 304	14 418	5637	1912	18 461	14 035	15 593	11 559	311	19 702	11 054	6233	1389
Thesis	10 919	33	3912	590	117	74	347	244	2891	35	35	702	1939
Conference	5741	718	162	447	159	375	626	505	240	587	468	439	1015
Other	649	20	112	305	2	3	62	2	93	2	0	1	47
Report	381	29	78	14	25	13	36	8	0	10	0	166	2
Unpublished	183	12	2	0	60	6	4	29	0	31	0	29	10
Patent	161	136	0	3	0	20	0	0	0	2	0	0	0
Total	861 839	65 929	134 000	67 000	67 003	65 836	65 400	67 000	63 347	66 726	66 998	65 740	66 860

EN: English; CN: Chinese; JP: Japanese; DE: German; ES: Spanish; FR: French; PT: Portuguese; KO: Korean; IT: Italian; PL: Polish; NL: Dutch; TR: Turkish

SERPs, document type identification becomes a rather complex task. By means of matching techniques to other data sources (WoS, CrossRef), and by applying a set of heuristics to extract more metadata from the websites that hosted the documents, the typology of 53.8% of the documents in the sample was ascertained (398 549). Table 4.18 shows the number of documents by typology and language.

If we focus only on the documents for which the typology could be identified, articles (65.3%) and books (30.2%) are the most frequent typologies, followed at a distance by doctoral theses and conference communications. However, we must bear in mind the limitations of the sample. The search strategy used (keyword-free searches which were only limited by year and language of publication), combined with GS's limitations (a maximum of 1000 results per query, sorted by GS's relevance ranking algorithm), result in a sample biased towards highly-cited documents, because the number of citations is the parameter that weighs the most in these types of queries [4.58].

In order to illustrate the bibliographic diversity in GS, Table 4.19 provides a list of the document types analyzed in the empirical studies that have addressed this issue.

Table 4.19 Document typologies found in GS

Bachelor's dissertations	Notes
Bibliographies	Presentation slides
Biographical items	Preprints
Blogs	Regulations
Book chapters	Reports
Book reviews	Research proposals
Books	Research reports
Civil service competitive examination reports	Reviews
Conference keynotes	Series
Conference paper proposals	Short survey
Conference papers	Student portfolios
Conference posters	Supplementary material
Conference presentations	Syllabi
Doctoral dissertation proposals	Term papers
Doctoral dissertations	Tweets
Doctoral qualifying examinations	Unpublished manuscripts
Editorials	Unpublished papers
Guidelines and clinical algorithms	Web documents
Interviews	Web pages
Journal articles	Working papers
Letters to the editor	Workshop papers
Master's theses	Yearbooks
Master's thesis proposals	

Clearly, GS's bibliographic wealth is due to the manner in which it indexes information: The search engine indexes any document that is hosted on the academic web, providing it meets certain technical and structural criteria. The consequence, at any rate, is that the presence of full text conference proceedings, book chapters, reports, patents, presentation slides (either from university courses, conferences, or other events), and especially monographs and doctoral theses make of GS a unique tool not only to find information, but also to find citation data that is not available anywhere else.

4.3.3 Growth Rate

Although Sect. 4.3.1 already presented some results as to this search engine's growth rate (even comparing it to Scopus and WoS), a sectional approach such as the one represented in Fig. 4.2 misses the main properties of GS, such as its dynamic nature. All content (both source documents and their citations, new or old) in GS is updated automatically.

As regards longitudinal analyses, *Harzing* [4.53, 59] studied the growth of citations to 20 Nobel Prize winners in 4 disciplines, detecting a growth of 4.6% from April 2012 to April 2013. A similar result (4.4%) was found by *Harzing* and *Alakangas* [4.54], where 146 senior researchers from the University of Melbourne were analyzed.

Retroactive growth (that is, the inclusion of documents published a long time ago) was addressed in part by *De Winter* et al. [4.52], who analyzed the relative difference between citation counts to a classic article up to mid-2005 measured in mid-2005 and citation counts up to mid-2005 measured in April 2013.

The speed with which GS indexes new source documents (and finds more citations to documents already in its document base) was addressed by *Moed* et al. [4.56]. The authors compute the indexing speed of GS for 12 journals in 6 different disciplines, and compare the results to those found in Scopus. Although there are differences among disciplines, and results are affected by the open-access policies of big publishers, the authors find that "the median difference in delay between GS and Scopus of indexing documents in Scopus-covered journals is about 2 months." The latest study on this issue to date was written by *Thelwall* and *Kousha* [4.60], which focuses on early citations to journal articles. They selected a sample of articles published in library and information science (LIS) journals between January 2016 and March 2017. The results in GS are compared to those in WoS, Scopus, and ResearchGate. The results in this study show that GS clearly outperforms all the other databases in terms of finding early citations,

Table 4.20 Speed of indexing for JASIST articles in GS

Article	Online publication	Scopus Index	WoS Index	GS				
				Index	Other version	Days since index	Online age	Index speed
1	20-Jan	Yes	No	No	Yes	–	–	–
2	24-Jan	Yes	No	No	Yes	–	–	–
3	27-Jan	Yes	No	Yes	No	56	58	2
4	21-Feb	No[a]	No	Yes	No	31	33	2
5	21-Feb	Yes	No	Yes	No	31	33	2
6	27-Feb	Yes	No	No	Yes	–	27	–
7	27-Feb	Yes	No	Yes	No	26	27	1
8	27-Feb	No[a]	No	Yes	No	26	27	1
9	27-Feb	Yes	No	Yes	No	26	27	1
10	27-Feb	Yes	No	Yes	Yes	N/A	27	–
11	07-Mar	Yes	No	Yes	No	17	19	2
12	07-Mar	Yes	No	No	Yes	–	19	–
13	13-Mar	Yes	No	No	Yes	–	13	–
14	20-Mar	No	No	Yes	No	N/A	6	–
15	20-Mar	No	No	No	Yes	–	6	–
16	20-Mar	No	No	Yes	No	3	6	3
17	20-Mar	No	No	Yes	No	3	6	3
18	20-Mar	No	No	Yes	Yes	N/A	6	–
19	20-Mar	No	No	Yes	No	3	6	3
20	20-Mar	No	No	No	Yes	–	6	–
21	20-Mar	No	No	No	Yes	–	6	–
22	25-Mar	No	No	No	Yes	N/A	1	–
23	25-Mar	No	No	No	No	–	1	–
24	25-Mar	No	No	No	No	–	1	–

[a] Not indexed in Scopus because they are book reviews
N/A: Not available

although ResearchGate's data is quickly becoming an interesting source for citation data as well.

What follows is a small-scale analysis that aims to illustrate this phenomenon, which has consequences not only for document searches, but on the speed with which citations are detected in the system. Articles accepted by the *Journal of the Association for Information Science and Technology* (JASIST) and made available as advance online publications between January 1st and March 25th, 2017 were identified, noting the specific date when they were made available online. Secondly, those articles were searched in GS, and, in the cases when they were found, the exact date of indexing was saved (this information is available when documents are sorted by date). Knowing these two dates, we were able to compute the speed of indexing (number of days since the article was first available online, until GS picked it up). Results are displayed in Table 4.20.

On the date of data collection (March 26th) GS had indexed 13 out of 24 of the articles analyzed from the publisher's website, although in 4 cases the date of indexing in GS was not available, because these documents were previously available in GS as preprints. What is more, out of the 11 articles GS had still not picked up from the publisher's website, 9 were available from other sources (mainly subject or institutional repositories). Only the two most recent articles (available on the publisher's website only 1 day before the analysis was carried out) were not available in GS in any form. As regards the indexing speed, it ranges from 1 to 3 days. It is worth taking into account that there is a ± 2 day margin of error, because we know the date of indexing but not the exact hour. According to these data, it seems that it only takes around 2 days for documents published in JASIST to be indexed in GS, although a larger sample would need to be analyzed to confirm this for sure.

If we compare these results to the coverage of these documents in other databases, we can observe that Scopus had indexed the documents that had been made available up to March 13th. Although the exact date of indexing in Scopus is unknown, it was necessarily below 13, a very respectable speed considering it is a controlled database. However, these documents are classified as *in press* in the platform, and Scopus does not compute citations to documents until they are formally published in a journal issue, something that can take months. WoS does not index documents until they have been formally published, and therefore does not cover any of the documents in the sample.

4.4 Google Scholar's Data for Scientometric Analyses

Lastly, this chapter would not be complete without addressing the limitations of this search engine as a source of data for bibliometric analyses.

It is important to differentiate between the limitations from which this platform suffers by design for a specific purpose, and the various kinds of errors that the search engine makes when it processes data from the academic web. Errors are deviations from the expected or normal functioning of the tool (like, for example, the existence of duplicate citations, versions of the same document that have not been merged, incorrect or incomplete attribution of authorship, etc.). Limitations, on the other hand, refer to the characteristics that can compromise the suitability of the tool for a specific purpose, especially if it is not the original purpose for which the tool was first developed (like, for example, using GS as a source of data for bibliometric analyses, instead of as a search tool).

4.4.1 Errors in Google Scholar

The studies that have been published on the topic of errors found in GS are rather disorganized and superficial. There are few empirical evidences, and they often lack proper systematic study backed by representative samples. Most of the time, only anecdotal evidence is presented, without addressing the important issue of the degree of pervasiveness of the errors (how often the errors occur throughout the document base). The results in these studies are difficult to summarize and compare, and they become obsolete very quickly, because GS is constantly being updated and introduces improvements to its algorithms regularly.

Following in the footsteps of the numerous and sharp studies carried out by *Jacsó* [4.12, 31, 54–69], below we present a taxonomy of the types of errors made by GS. We propose to divide errors into four broad groups (search-related errors, parsing-related errors, matching errors, and source-related errors), as described in Table 4.21. These errors can affect bibliographic records (authorship, source or year of publication, etc.) and citations themselves. A more in-depth discussion of the errors that can be found in GS was recently published by *Orduna-Malea* et al. [4.36].

4.4.2 Google Scholar Limitations

After describing the most common type of errors in the database, this section describes the main limitations for the use of GS as a source for bibliometric studies and research performance evaluation. To this end, we have prepared three descriptive sheets listing the limitations of GS search, GSC, and GSM, because although some limitations are present in all three products, some of them are particular to only some of them.

Each descriptive sheet contains several sections:

1. Coverage
2. Search and results interface

Table 4.21 Types of errors in GS

Type of error	Description
Search-related	Those related to the process of searching information
Parsing-related	Related to the process of identifying and extracting bibliographic information about documents from websites or full texts (including cited references)
Matching	Those related to the process of identifying different versions of a same document in order to remove duplicates
Source-related	Which affect the links that lead to the source in which the document has been found

Table 4.22 GS descriptive sheet

GS ***Coverage***
Lack of transparency in its coverage[a]: ● There is no public master list of the sources GS indexes (publishers, repositories, catalogues, bibliographic databases and repertoires, aggregators, ...). ● There is no public master list of journals indexed in the platform.
Non-scientific and non-academic documents are also occasionally covered: course syllabi, library guides, tweets, ...
There is no accurate method to estimate the size of GS.
Data is not stable. GS is dynamic and reflects the state of the academic web at a certain moment in time, ... The irregularity and unpredictability of GS's indexing speed may bias some bibliometric analyses if it is not taken into consideration.
Full text files that exceed 5 MB can be found on GS, but their full text will not be indexed (cited references will not be analyzed).
Easy to manipulate: Anyone can obtain a fully or partially fabricated document indexed in GS by uploading it to a university domain or public academic repository.

Table 4.22 (continued)

GS ***Search and results interface***
The advanced search form is limited to four search dimensions: keywords (with assisted Boolean operators, and the possibility to search only in the title of the document, or anywhere in the article), authors, source of publication (journal, conference, ...), and year of publication.
The number of records displayed in each results page is 10 by default. It can be increased to 20 on the settings page. In the past, however, it was possible to increase this number up to 100.
Only the first 1000 results for any query can be displayed. Similarly, even if a document has received more than 1000 citations, only the first 1000 can be displayed when clicking the *Cited by* link.
Results can only be sorted by relevance or by date of publication: • Relevance: This is the default method. Although the specific parameters that are taken into consideration for this sorting method have not been publicly disclosed, it has been found that the number of citations received by documents, as well as the language of the document in relation to the user's preferred language, both weigh heavily in the relevance sorting algorithm. • Publication date: limits the search to documents published in the current year.
There are only three result filtering options once a search has been made: • By document type, limited to three categories: case laws, patents, and articles. The latter category includes journal articles, books or book chapters, conference proceedings, technical reports, theses ... • By type of record: users are given the option to remove cited references (documents GS has only been able to find in the reference lists of other documents) from the search results. By default, cited references are included in the search results. • By year: it is possible to limit results to documents published in a given year, or a range of years.
It does not offer any features to analyze results or compute bibliometric indicators.
Quality of the data
In each search result, *authors* are displayed in the second line: below the title and next to the source of publication (usually a journal), the date of publication, and the name of the publisher or web domain where the document was found. The space allocated to the author data in this line is limited (usually between 30 and 40 characters), and, therefore, only the first three or four authors can be displayed, depending on the length of their names. In this line, authors are mentioned only by the initials of their first and second names and their surname. If these authors have created a public GSC profile and verified it with their institutional email, their name will contain a link to their public GSC profile. For more complete author data, users can click on the Cite button, and there, export the record to BibTeX (or another reference manager format). The BibTeX record will display the full name of up to 10 authors of the document.
No data regarding *institutional affiliation* of the authors of the documents is available (institution, country). Therefore, it is not possible to carry out studies on geographic and institutional production and collaboration.
There is no information available about the *language* in which documents are written, even though internally they must have this information, because users can choose to limit results to documents written in one or more of the following languages: simplified Chinese, traditional Chinese, Dutch, English, French, German, Italian, Japanese, Korean, Polish, Portuguese, Spanish, and Turkish.
The *typology of each document* is not clear (book, journal article, conference communication, thesis, report, ...). Only books are marked as such, usually when they have been found on Google Books.
Not all documents have an *abstract*
Author-supplied keywords are not available. The same happens with the descriptors used by the databases where the records are found.
The *list of cited references* in each article is not available either (even though they definitely have that information, since they need it to compute citations), making it difficult to carry out studies that require cited reference analysis.
Errors in the parsing routine can lead to numerous problems. There still is no conclusive study about the type and degree of occurrence of these errors, but among them we can find: • Poor bibliographic description of documents: incorrect or missing titles, authors, source of publication, date of publication, ... • Duplicate records, when GS is not able to match two or more records that actually refer to the same document. This can also lead to split citation counts, since some of the citations will be attributed to one of the versions of the document, and some to the other versions.
GS does not rely in any kind of controlled vocabulary for author names, journals, publishers, institutions, ... that facilitates the identification of the different name variants for these entities.
These last two limitations make it more difficult to carry out large-scale studies using GS data, since the data would have to go through important cleaning and normalization processes prior to the analysis.

Table 4.22 (continued)

GS
Data reuse and exporting capabilities
Users can copy citations for single records, in a variety of formats, by clicking on the *Cite* button below every record.
Records can be exported to reference managers manually one by one, also by clicking on the *Cite* button. The available formats are BibTeX, EndNote, RefMan, and RefWorks. Alternatively, users can also save records to the *My library* feature by clicking on the *Save* button (for which it is necessary to be logged in to a Gmail account). *My library* allows users to export up to 20 records in one go, in BibTeX, EndNote, RefMan, or CSV format. The abstract is never included as part of the exported records.
GS does not offer, nor is it planned to offer (at least in the near future), any kind of public application programming interface (API) to enable users to access and export data from GS in bulk.
A strict CAPTCHA system is in place to discourage users from making too many queries to the platform too quickly. Users (or bots) that go over a certain (undisclosed) number of queries in a certain time are asked to solve a CAPTCHA of some sort before they are able to continue their searches. Sometimes, if the system detects too many searches made from the same IP, that IP can get blocked temporarily.

[a] This limitation also affects GS citations and GSM

3. Quality of the data
4. Data reuse and exporting capabilities.

The information included has been taken from the authors' own observations and empirical tests. Due to length of the sheets, the sheet that describes the limitations of GS is given in Table 4.22, and the tables that list the limitations of GS citations and GSM can be found in the electronic supplementary material [4.14, see Appendix VI].

4.5 The Expanded Academic World of Google Scholar

The results of the analysis of the size, coverage, growth rate, and speed of indexing of GS fully justify considering this platform as a big data bibliographic tool (Fig. 4.10).

The empirical evidence described throughout this chapter allow us to affirm that GS is an all-inclusive tool, capable of bringing together not only the scientific world stricto sensu (that which is represented by WoS and Scopus as well), but the entire academic and professional world in a broad sense, thus providing a much broader picture of academic activity [4.37]. Its coverage is the most well balanced of the commonly used multidisciplinary databases in terms of countries of publication, languages (no English bias), and document typologies (not only scientific articles), something which is crucial when analyzing fields where it is common to use channels of communication other than journal articles published in English, such as disciplines

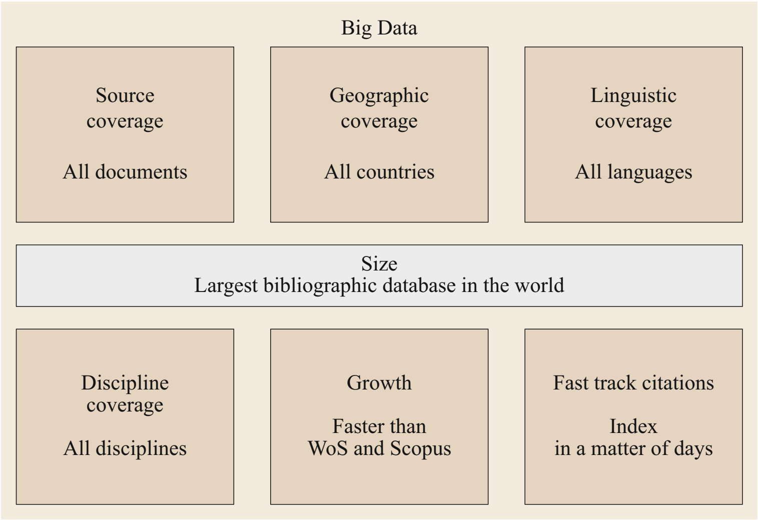


Fig. 4.10 GS, a big data bibliometric tool

in the arts, humanities, social sciences, and engineering. In this sense, Fig. 4.10 tries to convey the idea that GS covers all knowledge territories and all communities (scientific, educational, and professional), while WoS and Scopus only deal with scientific knowledge in the strict sense, and its communities.

The most important change over the previous paradigm, however, is GS's inclusion policy. WoS and Scopus have very exclusive and restrictive source inclusion policies. These sources are usually journals, which also place the prospective manuscripts researchers send to them under a rigorous evaluation processes (peer review). Contrary to this traditional model, GS works in a completely automated manner, without exercising any kind of selection process based on quality. Curated content from traditional journals and studies that have not gone through any kind of screening both coexist. GS automatically crawls, finds, and indexes any document that follows an academic structure and is hosted in an academic domain, even if it has not undergone any external quality control and is there only by decision of its authors. This breaks completely from the traditional controls to which all academic content had to be subjected prior to its public dissemination (peer review). For better or for worse, this is the distinguishing feature of GS: its ability to bring together reviewed and non-reviewed content, scientific and academic content (Fig. 4.11).

It should be pointed out that one of the main features shown in Fig. 4.11 is GS's nature as a receptacle. All the prestigious publications covered by WoS and Scopus are also covered by GS. When we observe the sources from which GS feeds, we can be sure, based on the empirical evidence on its size and coverage presented in the previous sections, that all the content covered by WoS and Scopus is also covered by GS. But of course, apart from the scientific elite, GS covers many other sources. We do not dispute that some of them may be of a lower quality, but others are of the same quality, if not higher, especially in certain disciplines (mainly doctoral theses, conference articles, working papers, and books).

Although it is true that GS covers sources that have not gone through a validation process like peer review (keynote conferences, syllabi, book reviews, technical reports, ...), these sources provide evidence of other kinds of impact beside the scientific impact, and could help put in a new light the work of researchers whose work is relevant to these communities and not the ones who publish in traditional databases.

Of course, the mixture of all these source documents (especially when considering that all of them are considered for computing citations) has been the object of important discussions in the bibliometric community [4.4]. To date, the main method to validate

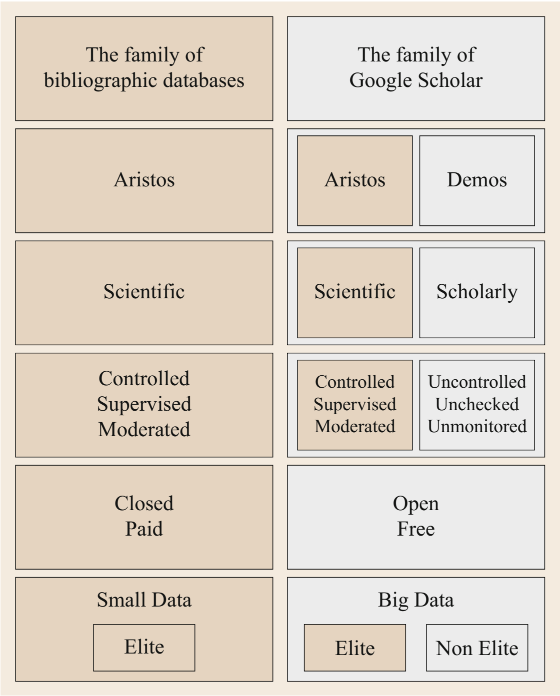


Fig. 4.11 GS versus traditional bibliographic databases

measurements made with data from GS has been to calculate their correlation with other well-established indicators. Many studies, which have been recently compiled by *Thelwall* and *Kousha* [4.70], have analyzed correlations between the number of citations according to GS and other databases (mainly WoS and Scopus), either at the level of journals or at the level of authors, as a way to evaluate its suitability as a source of data bibliometric studies.

The electronic supplementary material [4.14] offers a revised and updated list of these studies (see Appendixes II, III, and IV). Studies are grouped according to their unit of analysis (authors, citations, and average h-index). Although caution is advised for interpreting these results, because the nature of the samples are very different in terms of their size, time frames analyzed, and time when the studies were carried out, we can extract the following observations:

- The average correlation between GS and WoS (from 51 observations) is 0.76, and 0.81 when GS is compared to Scopus (28 observations).
- Out of all the studies, only in 2 are the correlations found to be below 0.50 (0.39 and 0.43 in Scopus, and 0.48 and 0.50 in WoS). All of them refer to correlations in the humanities and social sciences. On the other hand, there are numerous studies where the correlations found are above 0.9

(1 with GS/WoS comparisons and 7 with GS/Scopus comparisons).

- The highest correlations are found among STEM fields, and the lowest ones are usually found for fields in the humanities and social sciences.
- The differences between GS and the other databases seem to have increased in the more recent studies, which might indicate an even broader coverage in GS with respect to the other databases in recent years.
- No significative differences are found among studies with different units of analysis.

It also seems that multidisciplinary studies of international scope, and with very high sample sizes, achieve very high correlations, but that these become moderate by restricting the focus and emerging the intrinsic properties of each discipline. However, it should be borne in mind that correlations may have been calculated using different techniques (Pearson, Spearman, etc.), although values are reported independently, which may have a slight influence on the results.

To date, the largest sample studied for these purposes is the one used by *Martín-Martín* et al. [4.71], who used a sample of 64 000 highly-cited documents in GS, of which 51% were also covered by WoS and had at least one citation. Most of the documents covered by both databases were journal articles, and the rest were monographs, theses, and conference articles. The R-squared (Spearman) found for the number of citations received by documents covered by both databases was $R^2 = 0.73$.

For this chapter, we decided to replicate the previous study, using the sample of 861 843 highly-cited documents in 13 languages (see Appendix V in the electronic supplementary material for further details about this dataset) [4.14, 58]. Out of these documents, 69 279 (8%) of them were covered by WoS and had received at least one citation. The Spearman correlation between the number of citations according to GS and according to WoS for these documents was $R_s = 0.9$. Figure 4.12 presents a scatter plot based on these data.

As Fig. 4.12 shows, the correlation between the number of citations these documents have received according to the two databases is evident. Additionally, an important number of observations seem to have received many more citations from GS than from WoS. This means that, even though the correlation is very high in general terms, GS is usually able to find many more citations than WoS for the same documents.

4.6 Final Remarks

GS is a prodigious mine of academic information that covers all fields of knowledge. Thanks to GS we now have access to previously unexplored territories of knowledge which, even if only roughly, allow us to form a broader mental picture of academic activity. This platform sheds light where previously there was only darkness. GS can definitely help to open the academic Pandora's box [4.4]. Opening this box will bring to light document typologies (especially monographs, theses, reports, conference communications, and book chapters) from the scientific core and periphery, which were previously invisible and unaccessible. The bibliographic features of these documents will probably change certain axioms and prejudices of academic evaluation. Previously undervalued researchers, clearly harmed (or outright forgotten) by the policies of traditional databases, will come to light, because there will be evidence of the impact of their work. Conversely, it will also confirm the poor performance of other researchers, until now protected by the lack of proper tools to evaluate them.

Perhaps all this will finally lead to a redesign of certain assessment and promotion systems, funding programs, and even research policies and structures. The report *The metric tide* [4.72] already gives a glimpse of this changing trend from an institutional position, and not merely as an intellectual exercise advocated by a few and confined to research publications, with varying degrees of scientific impact, but without an actual practical impact.

That said, we cannot belittle the limitations of GS for bibliometric analyses. To begin with, the data exporting limitations (a maximum of 1000 results per query, and no easy way to export them). These obstacles are a hindrance when massive amounts of data are necessary for an analysis. The lack of an API in GS forces us to use third party applications (like Harzing's Publish or Perish) or download results manually. This results in very slow and costly data collection processes, which must be followed by a thorough cleaning of the raw data [4.63, 73, 74].

Additionally, GS does not provide vital information in its records, like the institutional affiliation of the authors, the language, and the document types. Not to mention the difficulties that normalizing the bibliographic information collected from so many varied sources entails. On the other hand, we do not believe that the errors from which some records suffer are

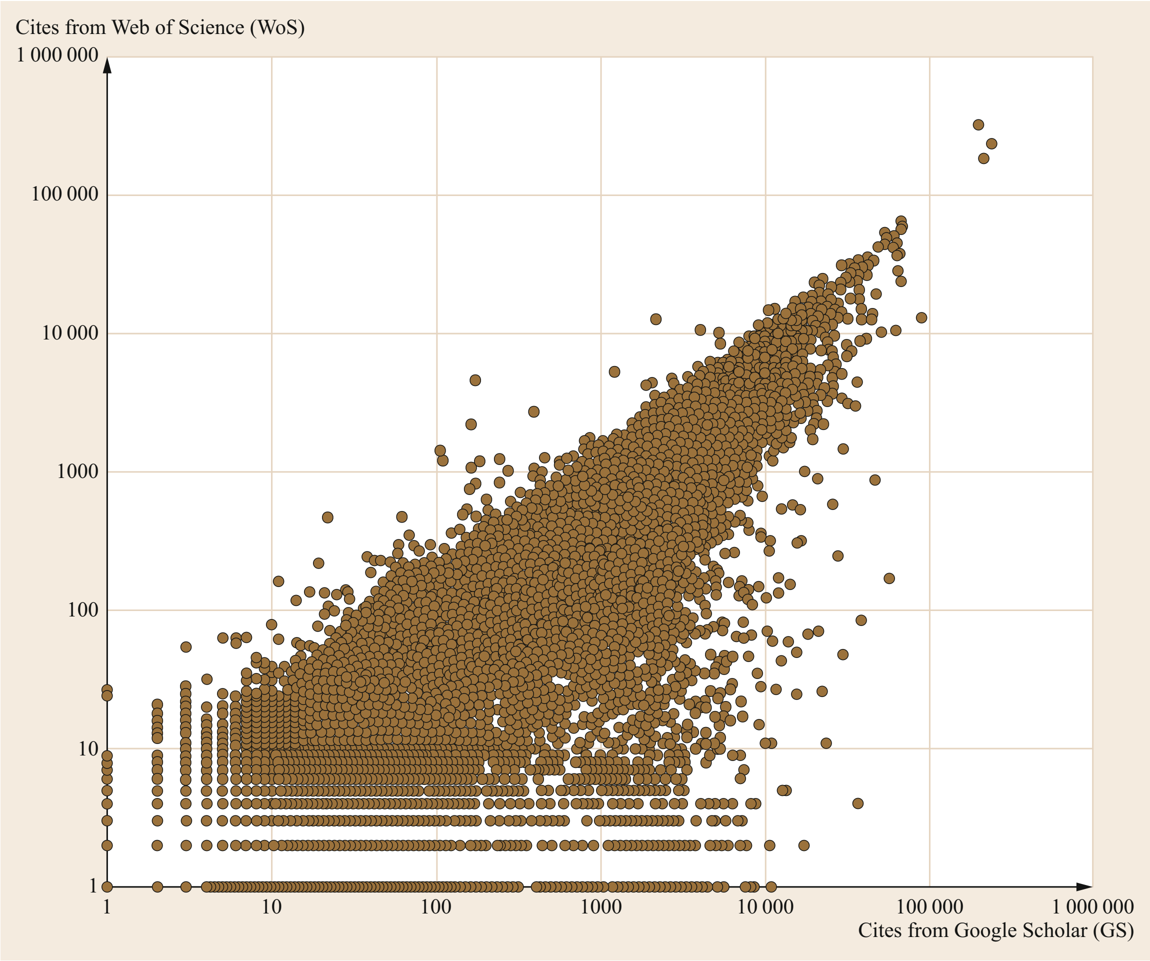


Fig. 4.12 Scatter plot of citations of highly-cited documents according to GS and WoS ($N = 69\,279$)

a major obstacle to the use of GS for bibliometric purposes. In a big data tool such as this, these errors are diluted and have no consequence on the big picture. These errors rarely affect individuals, journals, or other aggregates, as has been referred to by some studies already [4.75].

At any rate, the more dangerous limitations do not have to do with the methodological and technical problems previously discussed, but with the obscurity of the system, and, most of all, with the possibility of publication or citation manipulation, caused by the lack of quality control in the indexation of documents, which has been empirically proven by various studies [4.20, 76, 77]. Likewise, one of the main criticisms that is directed at GS is the lack of transparency, both regarding its coverage (what sources it indexes) and updating mechanisms, and regarding its algorithm for ranking results after a query.

Lastly, we hope that this study will help readers understand the inner workings of GS and become aware of its enormous potential. We always tried to offer empirical evidence of its strengths (*contra data non argumenta*), without forgetting about the important dangers that its abuse could lead to. We hope to make the scientific community question assumed truths of an academic world of which only the tip of the iceberg has been visible until now. Let us explore its depths, let us observe and describe these new landscapes, and then let us decide if we would rather remain on the surface.

References

4.1 E. Garfield, I.H. Sher, R.J. Torpie: *The Use of Citation Data in Writing the History of Science* (Institute for Scientific Information, Philadelphia 1964)
4.2 E. Garfield: Science citation index: A new dimension in indexing, Science **144**(3619), 649–654 (1964)
4.3 J.L. Ortega: *Academic Search Engines: A Quantitative Outlook* (Elsevier, Oxford 2014)
4.4 E. Orduna-Malea, A. Martín-Martín, J.M. Ayllón, E. Delgado López-Cózar: *La revolución Google Scholar: abriendo la caja de Pandora académica* (UNE, Granada 2016)
4.5 Google Scholar: My citation counts have gone down. Help!, https://scholar.google.com/intl/en/scholar/help.html#corrections (2017)
4.6 S. Baggett, M. Williams: Budgets stretched, staff stressed, usage stalled. . . Something's gotta give! In: *Proc. Charleston Library Conf.*, *Charleston*, ed. by B.R. Bernhardt, L.H. Hinds, K.P. Strauch (2012) pp. 324–327
4.7 B. Kramer, J. Bosman: *101 Innovations in Scholarly Communication–The Changing Research Workflow* (Universitait Utrecht, Utrecht 2015)
4.8 S. Brown, S.A. Swan: Researchers' use of academic libraries and their services. A report commissioned by the Research Information Network and the Consortium of Research Libraries (2007)
4.9 J. Mussell, R. Croft: Discovery layers and the distance student: Online search habits of students, J. Lib. Inf. Serv. Distance Learn. **7**(1/2), 18–39 (2013)
4.10 D. Nicholas, C. Boukacem-Zeghmouri, B. Rodríguez-Bravo, J. Xu, A. Watkinson, A. Abrizah, E. Herman, M. Świgoń: Where and how early career researchers find scholarly information, Learn. Publ. **30**(1), 19–29 (2017)
4.11 R. Van Noorden: Online collaboration: Scientists and the social network, Nature **512**(7513), 126–169 (2014)
4.12 P. Jacsó: Google Scholar Author Citation Tracker: Is it too little, too late?, Online Inf. Rev. **36**(1), 126–141 (2012)
4.13 E. Delgado López-Cózar, A. Cabezas-Clavijo: Google Scholar Metrics: An unreliable tool for assessing scientific journals, Prof. Inf. **21**(4), 419–427 (2012)
4.14 E. Delgado López-Cózar, E. Orduna-Malea, A. Martín-Martín: Supplementary material to 'Google Scholar as a data source for research assessment', https://doi.org/10.17605/OSF.IO/PQR53 (2017)
4.15 S.C. Bradford: Sources of information of specific subjects, Engineering **137**, 85–86 (1934)
4.16 Google Scholar: Google Scholar's inclusion guidelines for webmasters, https://scholar.google.com/intl/en/scholar/inclusion.html#indexing (2017)
4.17 Google Scholar: Google Scholar's About page, https://scholar.google.com/intl/en/scholar/about.html (2017)
4.18 Google Scholar: Google Scholar's search tips: Content coverage. What specific journals do you cover?, https://scholar.google.com/intl/en/scholar/help.html#coverage (2017)
4.19 E. Delgado López-Cózar, A. Cabezas-Clavijo: Ranking journals: Could Google Scholar Metrics be an alternative to journal citation reports and Scimago journal rank?, Learn. Publ. **26**(2), 101–113 (2013)
4.20 E. Delgado López-Cózar, N. Robinson-García: Repositories in Google Scholar Metrics or what is this document type doing in a place as such, Cybermetrics **16**(1) (2012)
4.21 A. Martín-Martín, J.M. Ayllón, E. Orduna-Malea, E. Delgado López-Cózar: Google Scholar Metrics 2014: A low cost bibliometric tool, EC3 Working Papers 17 (2014)
4.22 A. Martín-Martín, J.M. Ayllón, E. Orduna-Malea, E. Delgado López-Cózar: 2016 Google Scholar Metrics released: A matter of languages. . . and something else, EC3 Working Papers 22 (2016)
4.23 J. Bar-Ilan: Search engine results over time: A case study on search engine stability, Cybermetrics **2**(3) (1999)
4.24 R. Rousseau: Daily time series of common single word searches in Altavista and Northernlight, Cybermetrics **2**(3) (1999)
4.25 M. Thelwall: Quantitative comparisons of search engine results, J. Am. Soc. Inform. Sci. Technol. **59**(11), 1702–1710 (2008)
4.26 A. Uyar: Investigation of the accuracy of search engine hit counts, J. Inform. Sci. **35**(4), 469–480 (2009)
4.27 W. Koehler: Web page change and persistence–A four-year longitudinal study, J. Am. Soc. Inform. Sci. Technol. **53**(2), 162–171 (2002)
4.28 J.L. Ortega, I.F. Aguillo, J.A. Prieto: Longitudinal study of content and elements in the scientific web environment, J. Inform. Sci. **32**(4), 344–351 (2006)
4.29 N. Payne, M. Thelwall: Longitudinal trends in academic web links, J. Inform. Sci. **34**(1), 3–14 (2008)
4.30 A. Bosch, T. Bogers, T.M. Kunder: Estimating search engine index size variability: A 9-year longitudinal study, Scientometrics **107**(2), 839–856 (2016)
4.31 P. Jacsó: As we may search–comparison of major features of the Web of Science, Scopus, and Google Scholar, Curr. Sci. **89**(9), 1537–1547 (2005)
4.32 I.F. Aguillo: Is Google Scholar useful for bibliometrics? A webometric analysis, Scientometrics **91**(2), 343–351 (2012)
4.33 I.F. Aguillo: *Why Google Scholar? An Empirical Comparison of the Main Evaluation Tools* (Google Scholar Day, Madrid 2017)
4.34 E. Orduna-Malea, J.M. Ayllón, A. Martín-Martín, E. Delgado López-Cózar: Methods for estimating the size of Google Scholar, Scientometrics **104**(3), 931–949 (2015)
4.35 M. Khabsa, C.L. Giles: The number of scholarly documents on the public web, PLoS ONE **9**(5), e93949 (2014)
4.36 E. Orduna-Malea, A. Martín-Martín, E. Delgado López-Cózar: Google Scholar as a source for scholarly evaluation: A bibliographic review of database errors, Rev. Esp. Doc. Cient. **40**(4), e185 (2017)

4.37 A. Martín-Martín, E. Orduna-Malea, J.M. Ayllón, E. Delgado López-Cózar: A two-sided academic landscape: Snapshot of highly-cited documents in Google Scholar (1950–2013), Rev. Esp. Doc. Cient. **39**(4), e149 (2016)
4.38 H.R. Jamali, M. Nabavi: Open access and sources of full-text articles in Google Scholar in different subject fields, Scientometrics **105**(3), 1635–1651 (2015)
4.39 A. Verstak, A. Acharya: Identifying a primary version of a document, US Patent US8095876B1 (Washington DC 2013)
4.40 ResearchGate: About, https://www.researchgate.net/about (2017)
4.41 E. Orduna-Malea, E. Delgado López-Cózar: Google Scholar metrics evolution: An analysis according to languages, Scientometrics **98**(3), 2353–2367 (2014)
4.42 G.A. Barnett, C.J. Chung, H.W. Park: Uncovering transnational hyperlink patterns and web-mediated contents: A new approach based on cracking. com domain, Soc. Sci. Comput. Rev. **29**(3), 369–384 (2011)
4.43 E. Delgado López-Cózar, R. Repiso-Caballero: El impacto de las revistas de comunicación: Comparando Google Scholar metrics, Web of Science and Scopus, Comunicar **21**(41), 45–52 (2013)
4.44 L.M. Reina Leal, R. Repiso-Caballero, E. Delgado López-Cózar: H Index of scientific Nursing journals according to Google Scholar metrics (2007–2011), EC3 Reports 5 (2013)
4.45 E. Delgado López-Cózar, E. Orduna-Malea, D.M. Cartagena, E. Jiménez-Contreras, R. Ruiz-Pérez: Journal Scholar: Una alternativa internacional, gratuita y de libre acceso para medir el impacto de las revistas de Arte, Humanidades y Ciencias Sociales, EC3 Working Papers 5 (2012)
4.46 A. Martin-Martin, E. Orduna-Malea, A.W. Harzing, E. Delgado López-Cózar: Can we use Google Scholar to identify highly-cited documents?, J. Inform. **11**(1), 152–163 (2017)
4.47 E. Delgado López-Cózar, D.M. Cartagena, E. Jiménez-Contreras, R. Ruiz-Pérez: Índice H de las revistas españolas de Ciencias Sociales y Jurídicas según Google Scholar (2002–2011), EC3 Reports 4 (2013)
4.48 Grupo de Evaluación de la Ciencia y de la Comunicación Científica (EC3): IN-RECS database, http://ec3.ugr.es/in-recs (2017)
4.49 J.M. Ayllón, A. Martín-Martín, E. Orduna-Malea, E. Delgado López-Cózar: Índice H de las revistas científicas españolas según Google Scholar metrics (2011–2015), EC3 Reports 17 (2016)
4.50 X. Gu, K. Blackmore: Characterisation of academic journals in the digital age, Scientometrics **110**(3), 1333–1350 (2017)
4.51 K. Kousha, M. Thelwall: Google Scholar citations and Google Web/URL citations: A multi-discipline exploratory analysis, J. Am. Soc. Inform. Sci. Technol. **58**(7), 1055–1065 (2007)
4.52 J.D.F. De Winter, A.A. Zadpoor, D. Dodou: The expansion of Google Scholar versus Web of Science: A longitudinal study, Scientometrics **98**(2), 1547–1565 (2014)
4.53 A.W. Harzing: A preliminary test of Google Scholar as a source for citation data: A longitudinal study of Nobel Prize winners, Scientometrics **94**(3), 1057–1075 (2013)
4.54 A.W. Harzing, S. Alakangas: Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison, Scientometrics **106**(2), 787–804 (2016)
4.55 A.W. Harzing, S. Alakangas: Microsoft Academic: Is the phoenix getting wings?, Scientometrics **110**(1), 371–383 (2016)
4.56 H. Moed, J. Bar-Ilan, G. Halevi: A new methodology for comparing Google Scholar and Scopus, J. Inform. **10**(2), 533–551 (2016)
4.57 L. Bornmann, W. Marx, H. Schier, E. Rahm, A. Thor, H.D. Daniel: Convergent validity of bibliometric Google Scholar data in the field of chemistry citation counts for papers that were accepted by Angewandte Chemie International Edition or rejected but published elsewhere, using Google Scholar, Science Citation Index, Scopus, and Chemical Abstracts, J. Inform. **3**(1), 27–35 (2009)
4.58 A. Martín-Martín, E. Orduna-Malea, E. Delgado López-Cózar: Highly cited documents in Google Scholar by language and year of publication (January 2017), https://doi.org/10.6084/m9.figshare.5024465.v1 (2017)
4.59 A.W. Harzing: A longitudinal study of Google Scholar coverage between 2012 and 2013, Scientometrics **98**(1), 565–575 (2014)
4.60 M. Thelwall, K. Kousha: ResearchGate versus Google Scholar: Which finds more early citations?, Scientometrics **112**(2), 1125–1131 (2017)
4.61 P. Jacsó: Google Scholar: The pros and the cons, Online Inf. Rev. **29**(2), 208–214 (2005)
4.62 P. Jacsó: Deflated, inflated and phantom citation counts, Online Inf. Rev. **30**(3), 297–309 (2006)
4.63 P. Jacsó: Google Scholar revisited, Online Inf. Rev. **32**(1), 102–114 (2008)
4.64 P. Jacsó: Testing the calculation of a realistic h-index in Google Scholar, Scopus and Web of Science for F.W. Lancaster, Libr. Trends **56**(4), 784–815 (2008)
4.65 P. Jacsó: The pros and cons of computing the h-index using Google Scholar, Online Inf. Rev. **32**(3), 437–452 (2008)
4.66 P. Jacsó: Calculating the h-index and other bibliometric and scientometric indicators from Google Scholar with the publish or perish software, Online Inf. Rev. **33**(6), 1189–1200 (2009)
4.67 P. Jacsó: Google Scholar's Ghost Authors, Libr. J. **134**(18), 26–27 (2009)
4.68 P. Jacsó: Metadata mega mess in Google Scholar, Online Inf. Rev. **34**(1), 175–191 (2010)
4.69 P. Jacsó: Google Scholar duped and deduped–The aura of 'robometrics', Online Inf. Rev. **35**(1), 154–160 (2011)
4.70 M. Thelwall, K. Kousha: Web indicators for research evaluation. Part 1: Citations and links to academic articles from the Web, El profesional de la información **24**(5), 587–606 (2015)
4.71 A. Martín-Martín, E. Orduna-Malea, J.M. Ayllón, E. Delgado López-Cózar: Does Google Scholar contain all highly cited documents (1950–2013)?, EC3 Working Papers 19 (2014)

4.72 J. Wilsdon, L. Allen, E. Belfiore, P. Campbell, S. Curry, S. Hill: *The Metric Tide: Independent Review of the Role of Metrics in Research Assessment and Management* (HEFCE, UK 2015)
4.73 L.I. Meho, K. Yang: Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar, J. Am. Soc. Inform. Sci. Technol. **58**(13), 2105–2125 (2007)
4.74 J. Bar-Ilan: Citations to the “Introduction to informetrics” indexed by WOS, Scopus and Google Scholar, Scientometrics **82**(3), 495–506 (2010)
4.75 G. Doğan, İ. Şencan, Y. Tonta: Does dirty data affect google scholar citations?, Proc. Assoc. Inform. Sci. Technol. **53**(1), 1–4 (2016)
4.76 J. Beel, B. Gipp: Academic search engine spam and Google Scholar's resilience against it, J. Electron. Publ. (2010), https://doi.org/10.3998/3336451.0013.305
4.77 C. Labbé, D. Labbé: Duplicate and fake publications in the scientific literature: How many SCIgen papers in computer science?, Scientometrics **94**(1), 379–396 (2013)

Emilio Delgado López-Cózar

Información y Documentación
Universidad de Granada
Granada, Spain
edelgado@ugr.es

Emilio Delgado López-Cózar is a Professor at the Facultad de Comunicación y Documentación of the Universidad de Granada. He holds a PhD in Library & Information Science. His research interests are focused on the analysis of the new sources of information and scientific evaluation (e.g., search engines like Google Scholar) and of the new metrics used to measure the visibility and the impact of scientific activity (Altmetrics).

Enrique Orduna-Malea

Audiovisual Communication,
Documentation, and History of Art
Universitat Politècnica de València
Valencia, Spain
enorma@upv.es

Enrique Orduna-Malea is Assistant Professor at the Polytechnic University of Valencia. He holds a PhD in Webometrics, a Master Degree in Management of Multichannel Contents, and MA in Library & Information Science. His research interests focus on the study of contents and users in the Web and the design of methods and models of analysis. He has co-authored 3 books, and 40+ articles included in JCR (Journal Citation Reports).

Alberto Martín Martín

Información y Documentación
Universidad de Granada
Granada, Spain
albertomartin@ugr.es

Alberto Martín Martín is a fourth-year PhD candidate at the University of Granada, thanks to a Doctoral Fellowship granted by the Ministerio de Educación, Cultura y Deporte in Spain. For his thesis project, he is studying the potential of the data available in Google Scholar for bibliometric analyses.

5. Disentangling Gold Open Access

Daniel Torres-Salinas, Nicolas Robinson-García, Henk F. Moed

This chapter focuses on the analysis of current publication trends in gold Open Access (OA). The purpose of the chapter is to develop a full understanding of country patterns, OA journal characteristics and citation differences between gold OA and non-gold OA publications. For this, we will first review current literature regarding Open Access and its ostensible citation advantage. Starting with a chronological perspective we will describe its development, how different countries are promoting OA publishing, and its effects on the journal publishing industry. We will deepen the analysis by investigating the research output produced by different units of analysis. First, we will focus on the production of countries with a special emphasis on citation and disciplinary differences. A point of interest will be identification of national idiosyncrasies and the relation between OA publication and research of local interest. This will lead to our second unit of analysis, OA journals indexed in Web of Science. Here we will focus on journal characteristics and publisher types to clearly identify factors which may affect citation differences between OA and traditional journals which may not necessarily be derived from the OA factor. Gold OA publishing, as opposed to green OA, is being encouraged in many countries. This chapter aims at fully understanding how it affects researchers' publication patterns and whether it ensures an alleged citation advantage as opposed to non-gold OA publications.

5.1 Open Access and Scholarly Communication

Almost 30 years have gone by since the emergence of ArXiV, the revolutionary open access launched in a pre-Internet era in the early 1990s [5.1]. This event established the first landmark of the open access (OA) movement, a revolution in the way scholarly works are disseminated which would not have been possible without the technological advancements that preceded it. The spread of OA has always been surrounded by controversy with regard to its motivations, and its effects and the benefits for those providing open access. Among others, *Kurtz* and *Brody* [5.2] have pointed out the *increasing access* to scientific information, identifying OA as a natural step in today's ever more rapid communication processes. Contrarily, *Beall* [5.3] sees OA publishing as a threat to the scholarly communication system, and argues that "authors, rather than libraries, are the customers of open access publishers, so a powerful incentive to maintain quality has been removed".

The current expansion of OA can be explained partially by the serial breakdown [5.4], a scholarly publishing crisis derived from the shift to online scientific publishing and the concentration of journals among a few publishers, who obliged libraries to subscribe to fixed collections of journals at unsustainable prices through *big deals* [5.5], limiting access to scientific literature. However, OA publishing is now handled by a few publishers who negotiate country-wide licenses for article processing charges (APC) through deals which share many similarities with the *big deal* subscription access agreements [5.6]. *PLOS One*, *Scientific Reports* and *Nature Communications*, the three most

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_5

prolific mega-journals, already represent 62.2% of all gold OA publications between 2012 and 2016 (data extracted from Clarivate Analytics InCites). In the same vein, PLOS itself generates more than $40 million annually (https://www.plos.org/financial-overview). It is safe to state therefore, that OA is no longer just an ideological movement, but also a multimillion business.

The OA movement took shape in the Budapest Open Access Initiative (http://www.budapestopenaccessinitiative.org/) in 2002 seeking to unite:

> [A]n old tradition and a new technology [which had] converged to make possible an unprecedented public good .... [T]he willingness of scientists and scholars to publish the fruits of their research in scholarly journals without payment, for the sake of inquiry and knowledge.

Still, traditional publishers have constantly resisted modifying their subscription-based business model, leading to boycotts from scientists [5.7] and calls to defy publishers' copyright privileges [5.8]. While some concessions have been made, and most journals nowadays allow preprint versions of papers to be uploaded in repositories, such defiance to break paywalls still remains, either through legal [5.9] or illegal means [5.10].

Many studies and reviews have been devoted to defining the characteristics and implications of OA [5.11–13], the state of OA [5.14–17], how much of the scientific literature OA represents [5.18], benefits derived from it [5.19, 20] or its relation with citation impact [5.21–24] and other related topics. Still, there are many questions and misunderstandings that need to be addressed to better comprehend and assess the mechanisms that are being put in place to make OA possible.

In this chapter, we focus on gold OA, that is, publications from OA journals, and analyze differences in production and impact between countries by using normalized citation scores to better characterize the phenomenon of OA publishing. Long gone is the debate questioning the viability of an OA publishing model [5.25]. Many OA journals are now well-known and well-established journals, and authors have accepted the APC model and are willing to pay journal publication fees. However, there are important sectors of the scientific community who still raise concerns as to the quality of these journals and the potential threats they pose to the scientific communication system [5.26–28].

5.2 What is Open Access?

The Budapest Open Access Initiative (BOAI) was the first to coin the term *open access* [5.19]. In their founding document, they offered the following definition of OA:

> [Research literature which is] free[ly] availab[le] on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution and the only role for copyright in this domain should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited.

The Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (BDOA) of 2003 (https://openaccess.mpg.de/Berlin-Declaration) added further specifications to this definition. First, it defined research literature as:

> Original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material.

Second, it established that OA contributions must comply with two conditions:

1. The concession of copyrights to access, use, distribute and modify freely and worldwide such documents as long as the author(s) is acknowledged.
2. That the document is deposited in an online repository which complies with certain technical standards.

But the implementation of OA led to situations which were not originally contemplated by these definitions. One of them has to do with the appearance of multiple versions of the same document [5.29]. As a document can be uploaded to a repository and also submitted to a journal, various versions of a paper can be available at the same time (pre-submitted manuscript, peer-reviewed manuscript and journal formatted version). This situation can create a level of uncertainty, as there could be substantial differences between these different versions. It must be noted that in many cases the version uploaded to a repository is a so-called author copy of the manuscript published in the journal, i. e., the version accepted for publication by the journal, but not subjected to the copy-editing process conducted by the publisher. This practice ensures that the scientific contents of the preprint and published version are identical.

Also, the use of relaxed notions of OA influence the perception researchers have as to what OA is and what it is not. There is a tendency to consider OA and free access as synonyms. OA was formally defined more than ten years after the practice had begun. Issues such as copyrights or how to make documents accessible were not even considered at this early stage [5.25]. It is plausible to speculate that the retroactive definition of OA has helped resolve this misconception of OA. For instance, a recent report commissioned by the European Commission claimed that more than half of the publications were in OA, but defined OA as "freely available online to all (no money had to be paid, no registration to a service or website had to be made)" [5.18]. This was later noted by a news story published in Nature which had to emphasize that [5.30]:

> Although free to read, [articles] may not meet formal definitions of open access because, for example, they do not include details on whether readers can freely reuse the material.

Although free access can still be perceived as positive, the fact that authors are uploading their publications to private corporations such as ResearchGate, Mendeley, figshare or Academia instead of OA repositories, raises concerns as to the future sustainability of the OA movement [5.6].

Originally, two routes to OA were envisioned to provide a middle ground that offers room to new business models while promoting sustainable and universal access to scientific literature. These are known as the green and golden routes [5.16]: two non-mutually exclusive models to reach OA while conceding space to journals to make a profit. The green route designates scientists as those to be held responsible for permitting OA to their publications. They are expected to upload their works to repositories maintained by university libraries following the model set by *Ginsparg* and ArXiV [5.14, 31]. In principle, this solution leaves the journal as an accessory element which ceases to be at the core of the scientific communication system. Still, journals influence authors' decisions on making their work accessible, as they hold the copyrights of the manuscripts they publish [5.19].

The golden route maintains journals as the core of academic communication. Journals are the ones which should provide OA. This means abandoning the subscription-based business model. *Laakso* et al. [5.17] describe three types of OA journals: direct OA, delayed OA or hybrid OA. Direct OA journals are those which offer their full contents in OA. While most of these publishers adopt an author-pays model, including article processing charges (APC), this is not a prerequisite for direct OA journals and, in fact, many journals are maintained by public institutions without incorporating any fee for authors. Delayed OA journals, on the other hand, maintain a subscription-based model but offer OA to all their contents after a given period. Hybrid OA journals are the most restrictive of the three types. In fact, many argue that they are not true OA journals, as they provide OA only to those publications for which the authors have paid an OA clause, the rest of their contents remaining behind paywalls.

Since these two routes were defined, other types of OA have been described in the literature. We must note that some of these denominations are controversial and are not even considered by some authors as truly OA, as they do not fit the general definition provided by the Budapest Open Access Initiative. For instance, *Suber* [5.32] makes the distinction between Gratis and Libre OA. He defines the former as that which offers rights only to *read* articles, whereas the latter extends rights to reuse articles. Another proposed type of OA is that named as Black OA [5.33], which is defined as that in which articles are illegally distributed through pirate sites such as Sci-Hub and LibGen, something considered by many as not real OA. Finally, it is worth discussing bronze OA [5.34]. While not truly OA, articles falling under this category are those which are freely distributed by publishers through their websites, without including any type of open access license. Articles falling under this category respond to forms of delayed OA (that is, journals making their archives accessible), from journals not listed in the *Directory of Open Access Journals* (DOAJ), or articles which are temporarily offered freely by journals to promote their contents.

One of the main concerns of OA advocates has been to learn how much of the scientific literature is already in open access and how this number is growing in respect to the overall growth of scientific literature. *Björk* et al. [5.15] established that 20.4% (up to 24% of articles in hybrid journals are included) of the literature was already OA in 2008, 8.5% being gold OA. *Gargouri* et al. [5.35] found similar figures when analyzing its growth and differences by discipline. They reported that an average of 24% of the literature was in OA between 2005 and 2010. Contrarily to Björk and his team, they found out that only 2% of the OA literature was gold OA, with biomedical research being the field with the highest share of gold OA (8%). A more recent study [5.34] reported that up to 28% of the literature was accessible via OA, bronze OA being the most common form of OA (16.2%) followed by green OA (4.8%), hybrid OA (3.6%) and, lastly, gold OA (3.2%).

First studies analyzing publication and citation trends in OA tended to focus on green OA [5.22–24], as in many cases the authors of these studies were themselves advocating for green OA [5.36]. But journals' embargo requirements and OA release delays have pushed others to considered gold OA as a more promising venue [5.37], leading to new studies focused specifically on OA journals. Since the beginning of the 2000s the number of OA journals has grown exponentially, going from fewer than 750 journals in 2000 to more than 6500 in 2011 [5.38]. The types of journals have also diversified, and we can now differentiate between OA journals with or without an APC model, born versus converted OA journals and between small and mega-journals.

5.3 Disentangling Gold Open Access

One of the first issues studies on OA journals encountered was the difficulty to find a comprehensive list of all OA journals [5.29, 36]. Recently, both Scopus and Web of Science (WoS) have provided the ability to identify OA journals within their databases. Data and figures displayed in this chapter are retrieved from Web of Science InCites and hence some discrepancies could be found with OA journal lists provided elsewhere. According to Clarivate, OA journals in InCites are based on the DOAJ open access status [5.39]. This must be considered when interpreting the findings displayed.

The success of OA publishing is intimately related to the expansion of digital distribution and the consequent abandonment of printed materials. According to *Björk* and *Solomon* [5.40] we can distinguish four waves that took place almost simultaneously in the latter half of the 1990s and the beginning of the new century. The first wave is characterized by the launch of new OA journals by individual scientists, exemplified by the *Journal of Medical Internet Research*, currently a world leading journal in its field. The second wave consists of the adoption of an OA model by established subscription journals. Here a pioneering journal was British Journal of Medicine (BMJ), first offering OA in 2000. A more ambitious initiative, in the sense of the magnitude of journals converting to OA, is the launch of Scielo in 1997, a publicly subsidized South American portal, which gives OA to hundreds of journals at no costs to publishers.

A game changer in OA publishing was the launch of *BioMed Central* (BMC, currently owned by Springer) and *PLOS* in 1998 and 2000 respectively. These two OA-born journals are the first ones to adopt an APC model, envisioned by BMC founder, Vitek Tracz. The fourth and last wave is that representing the reaction of established publishing firms to OA by proposing the hybrid model. That is, subscription-based publishers implementing an OA option at the level of individual articles, such as Springer's *Open Choice*, thus making journals hybrid in terms of the financing model.

The APC model is seen by many as the most sustainable solution for maintaining journals as a profitable business while ensuring universal access to scientific literature. Supranational and national organizations, such as the European Union, the US National Institutes of Health or the British Wellcome Trust, include OA policy mandates ensuring that all publicly funded research findings are preserved and publicly accessible. In most cases, they promote both the gold and green routes, considering that the APC model allows "savings of up to 30%" [5.39] compared with journal subscriptions. While it can be claimed that gold OA ensures the publication of peer-reviewed papers and hence the credibility of research, others have questioned if an author-pays model really ensures the quality of the work published [5.28]. As authors become essential to the business model of these journals, quality levels may drop or cease to exist, leaving room for *predatory journals* which will publish anything as long as profits keep rising [5.3]. This phenomenon, which is not directly related with gold OA but with the APC model, has caused many researchers to perceive gold OA publications as research of a lesser quality [5.26].

In the late 2000s a debate emerged in the literature with regard to a perceived OA citation advantage [5.22–24, 41, 42]. Green OA advocates promoted the idea that researchers who made their publications OA had a citation advantage, as opposed to those who did not, due to the higher accessibility and visibility of their work. While this perception was confirmed by most studies [5.43], some authors pointed out that this perception could be more of an acceleration of the number of citations rather than an advantage due to an *early view* bias that favored OA papers [5.24]. In the case of gold OA the story is rather the opposite: "open access has multiplied that underclass of journals, and the number of papers they publish" [5.28]. The pernicious effect of *predatory publishers* and the low barriers set by many OA peer-reviewed journals have led opponents of OA to state that "[t]he open-access movement has been a blessing to anyone who has unscientific ideas and wants to get these ideas into print" [5.27].

Björk and *Solomon* [5.40] argue that OA journals are not of a lesser quality than traditional ones, but are younger. In fact, OA journals in fields such as biomedicine, or those with an APC model indexed in Web of Science, have citation rates similar to those of subscription-based journals [5.44]. This is also corroborated by *Gumpenberger* et al. [5.36], who indicate that there is an "overall positive impact trend for top Gold Open Access journals".

5.3.1 Gold OA Output and Impact of Countries and Scientific Fields

In this section, we will examine such arguments by analyzing the most recent trends in OA publishing for the period 2007–2016. In contrast to the study by *Björk* and *Solomon* [5.40], we do not distinguish between OA funded by article processing charges (APC) and other, mainly subsidized, forms of OA. In addition, rather than categorizing journals according to the *country of the publisher*, and calculating journal impact factor-like citation rates for a journal as a whole, the current study also presents analyzes on the *article production of countries in OA journals*, as expressed in the affiliations of publishing authors, and so called *category normalized citation rates*, comparing the citation impact of an entity—e. g., the total collection of articles published in OA papers, or the OA articles published by authors from a particular country—to the world citation average in the subject field in which the entity is active.

Table 5.1A provides insight into the global development of gold OA publishing. The table shows a steady increase of this percentage over the years, from 3.24% in 2007 to 12.09% in 2016. For the period 2012–2016 this percentage amounts to 10.5%. The last column gives the category normalized citation impact of the OA articles published in the various years, correcting for differences not only in citation practices between subject fields, but also between document types and publication years. A value of 1.0 means that gold OA articles are cited on average as frequently as an average article (either gold OA or non-gold OA). The

Table 5.1 Overview of gold OA publications: Evolution and distribution by research fields. 2007–2016 period

A. Evolution of gold OA				
Publication year	**Articles WoS**	**OA Articles WoS**	**% OA Articles WoS**	**Category normalized citation impact**
2007	1 071 782	34 752	3.24	0.81
2008	1 158 136	50 616	4.37	0.76
2009	1 228 957	61 143	4.98	0.79
2010	1 284 685	73 543	5.72	0.81
2011	1 373 833	93 273	6.79	0.84
Total 2007–2011	**5 045 611**	**278 575**	**5.52**	**0.81**
2012	1 414 483	114 252	8.08	0.86
2013	1 484 570	143 753	9.68	0.89
2014	1 527 771	166 465	10.90	0.88
2015	1 555 307	180 337	11.59	0.83
2016	1 466 589	177 251	12.09	0.73
Total 2012–2016	**7 448 720**	**782 058**	**10.50**	**0.83**
B. OECD research fields gold OA				
OECD research field	**Articles WoS**	**OA Articles WoS**	**% OA Articles WoS**	**Category normalized citation impact**
Natural sciences	3 579 626	367 004	10.25	0.97
Engineering	1 820 952	123 131	6.76	0.60
Health sciences	2 238 476	319 198	14.26	0.79
Agricultural sciences	353 675	39 092	11.05	0.46
Social sciences	835 437	34 917	4.18	0.84
Humanities	347 368	8660	2.49	0.54

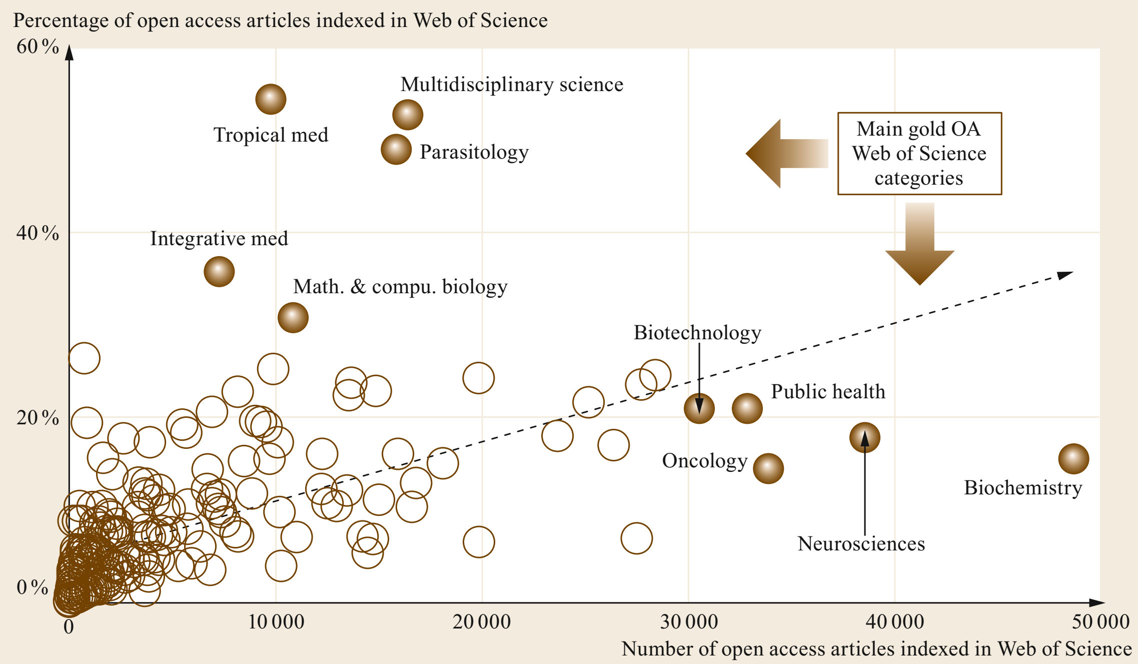
Technical Note. Dataset: InCites dataset; schema: OECD; document type: article; time period: 2007–2016
InCites dataset updated 13.05.2017. Includes Web of Science™ content indexed through 31.03.2017

category normalized citation impact of gold OA articles ranges over the years between 0.73 and 0.89. There is no clear trend in the data. Aggregating data into two 5 year periods, the scores in the two periods are statistically similar: 0.81 versus 0.83. In short, gold OA articles are cited some 15% less often than an average article (either gold OA or non-gold OA article).

Table 5.1B presents a breakdown of the gold OA articles and their impact by scientific field. 41.1% of OA articles are published in journals assigned to the natural sciences discipline. Relative to the total number of articles in this subject field, the share of gold OA articles is 10.25%, which is near the overall average of 10.5 indicated in Table 5.1. This is also the field with the largest normalized citation score (0.97), which is somewhat higher than the value of 0.83 obtained for the total set of all gold OA articles in the database from all disciplines.

Figure 5.1 delves into these areas by showing the total number of gold OA publications by WoS subject category and the percentage which gold OA represents within each subject category. Biochemistry is the field with the highest number of gold OA publications, followed by neurosciences, oncology and public health. However, it is in the subject categories of multidisciplinary sciences, tropical medicine and parasitology where gold OA represents half of the overall number of publications. The explanation of such a large share in the multidisciplinary fields is due the presence of OA mega-journals such as *PLOS One* and *Scientific Reports*. In the case of *tropical medicine and parasitology*, journals from big OA publishers are still present (e. g., *PLOS Neglected Tropical Diseases*, *PLOS Pathogens* or *Malaria Journal*, which belongs to BMC), but there are also national journals which do not publish in the English language. Interestingly, these are mostly Brazilian journals such as *Memorias Do Instituto Oswaldo Cruz*, *Revista da Sociedade Brasileira de Medicina Tropical* or *Revista Brasileira de Parasitologia Veterinaria*. This is reasonable when considering that these subject categories have a heavy local component and that South America and Brazil, in particular, are regions where these subject categories are more relevant in comparison with regions such as North America and Europe.

The ranking of countries according to their number of gold OA publications differs substantially from that based on total publication output (Fig. 5.2). Some top countries in terms of gold OA make a much smaller contribution to the total global output. For instance, Brazil, which occupies the 14th position of most pro-



Technical Note Dataset: InCites dataset; schema: OECD; document type: article; time period: 2007–2016
InCites dataset updated 13.05.2017. Includes Web of Science™ content indexed through 31.03-2017.

Fig. 5.1 Total number and share of gold OA publications by WoS subject category. 2012–2016 period ductive countries within the 2007–2016 period according to the Web of Science, is actually the fourth country in terms of the absolute number of OA publications. Similarly, Spain, the ninth country with the largest number of publications within the same period, goes up to the seventh position when one considers only gold OA publications. Another example is Mexico, which falls from the 23rd position according to their contribution to OA publishing, to the 31st position when looking at their global figures.

The case of Brazil is particularly interesting, as it is by far the country with the largest share of OA publications from its overall output (almost 30%), a consequence of the gold OA proactive policy undertaken by the Brazilian government through the promotion of the SciELO platform, an initiative followed by other Latin American and Caribbean countries, which provides OA to journals from these countries [5.45]. In the case of Spain, the share of OA publications based on its overall output is not as large as that of Brazil, but, as noted elsewhere [5.46], Spain has increased its share of gold OA output at a higher rate than the world average over the last decade.

While for all countries the category normalized impact of OA publications is lower than the overall value of their normalized impact, the gap between these two figures differs greatly by country. Differences in the case of the United States, United Kingdom, Japan, Switzerland and Austria are almost non-existent. This is not the case for China, Brazil, Spain, France or Canada, for which the impact of gold OA publications is substantially lower than their overall impact level. This difference can be partly explained by the disciplinary profiles of these countries.

Figure 5.3 shows the five most productive countries of gold OA publications by six scientific fields as defined by the Organization for Economic Co-operation and Development (OECD). Colors represent the value of the category normalized citation impact of each country-field combination. As observed, countries with higher normalized impact tend to publish most of their output in the medical and health sciences, the natural sciences, and social sciences. In these fields, most of the countries in the top five positions reach an overall impact equal to or above the world average (category normalized citation impact ≥ 1.0).

This is not the case in agricultural sciences, engineering and technology and humanities. All countries—except the USA in engineering and technology and in the humanities—have lower values of normalized impact than the world average. Also, these are the only fields where the United States is not the largest con-

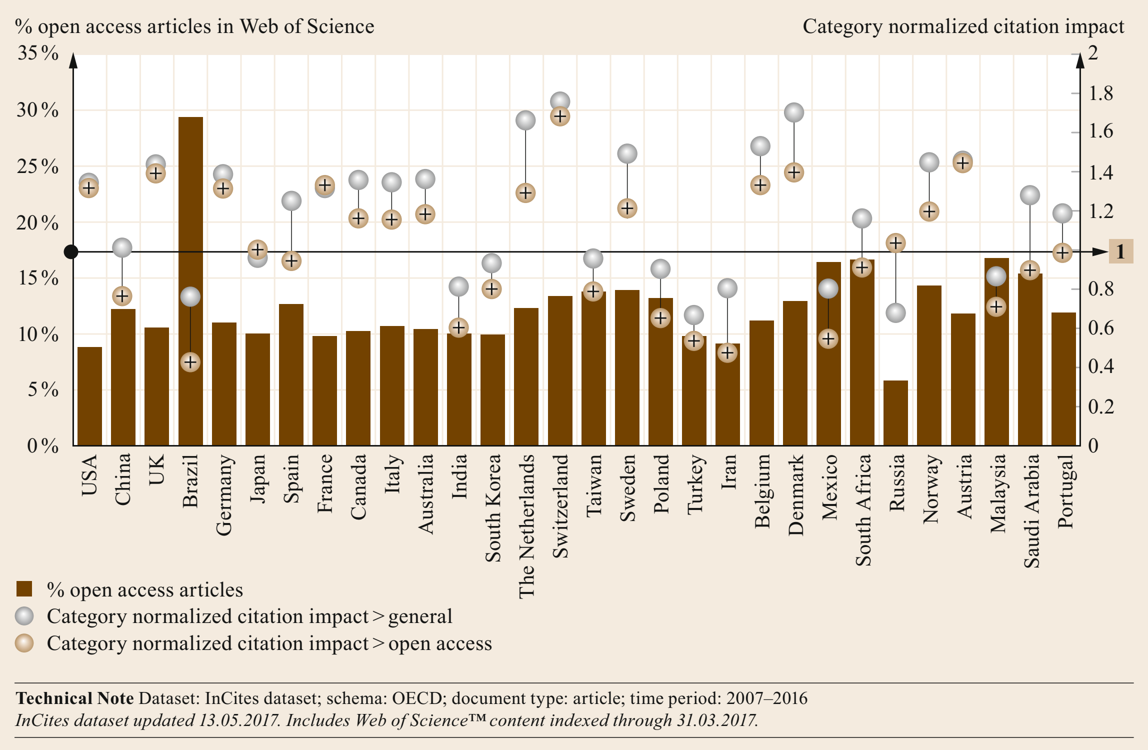


Fig. 5.2 Status of gold OA publications by country. 2012–2016 period. Data retrieved from Web of Science

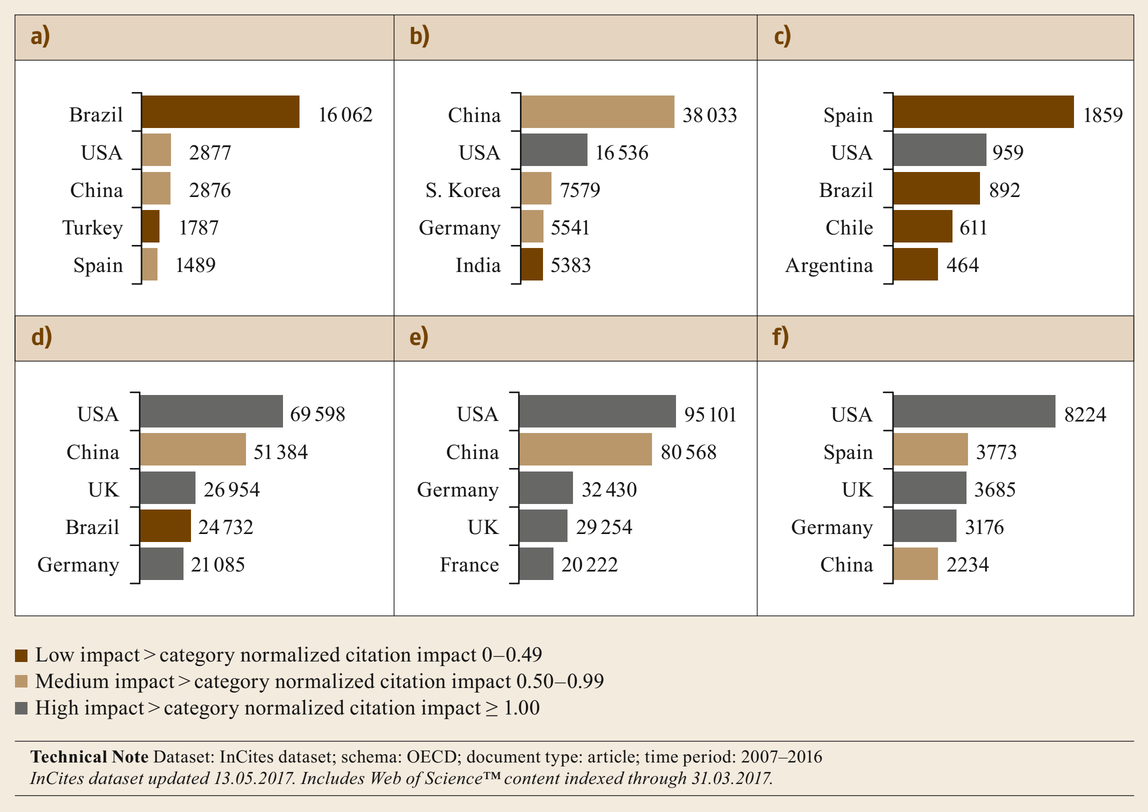


Fig. 5.3a–f Top five most productive countries in open access journals by field: **(a)** agricultural sciences; **(b)** engineering and technology; **(c)** humanities; **(d)** medical and health sciences; **(e)** natural sciences; **(f)** social sciences

tributor. Brazil, for instance, is the country with by far the largest number of publications in agricultural sciences. A closer examination in this field shows that six of the top ten journals producing most of the publications are Brazilian, all of which exhibit low impact factor values under 0.6. China is number one in the case of engineering and technology, and Spain in the case of the humanities. It is also worth noticing the asymmetry of the share of publications in Fig. 5.3a,b, where there is a large difference between the number of publications of the first country with respect to the rest.

These figures suggest that the differences in impact between gold OA papers and a country's total publication output, shown in Fig. 5.2, are largely due to the gold OA disciplinary profile of each country and the characteristics of OA journals in these different fields. This idea is reinforced by *McVeigh* [5.47], who indicated that high impact OA journals are unevenly distributed among countries and research fields. Furthermore, the large presence of Central and South American journals [5.48] could explain the large proportion of OA publications with low impact from countries such as Brazil, Spain, Chile or Argentina. Considering that there is an English-language positive bias and a Portuguese as well as Spanish-language negative bias in the citation impact scores of OA journals [5.49], it is appropriate to conclude that [5.46]:

> [T]he reasons behind [the] poor performance [of these countries] may be due to the national factor as well as to the large share of gold OA papers [compared to that of other] countries.

5.3.2 Characterizing OA Journals by Country and Field

Our next analysis focuses on the country of origin of gold OA journals. The purpose is to examine the characteristics of OA journals and illustrate how the relation between gold OA impact, countries and fields is intimately related or closely linked with the presence of national OA journals from these countries and fields. The issue between countries' output and the country

Table 5.2 Differences between Open Access journals and subscription-based journals. 2007–2016 period. Data retrieved from Web of Science

	Journals	WoS docs	Avg impact factor	St dev impact factor	Avg CNCI	St dev CNCI
Non-open access	10 194	6 554 091	1.96	3.02	0.85	1.07
Open access	1025	844 036	1.94	2.22	0.62	0.59
Total general	11 219	7 398 127	1.96	2.95	0.83	1.04

Technical Note. Dataset: InCites dataset; schema: OECD; document type: article; time period: 2007–2016
InCites dataset updated 13.05.2017. Includes Web of Science™ content indexed through 31.03.2017

of origin of journals has been widely examined when studying the phenomenon of *predatory journals* in OA publishing. *Bohannon* [5.28] published a controversial experiment in Science magazine where he submitted hundreds of fake manuscripts to journals indexed in the *Directory of Open Access Journals* (DOAJ) and Bealls' List of Predatory Journals [5.3]. His experiment criticized the lack of peer review of many of the journals to which he submitted his bogus papers which were, in many cases, accepted. This paper was seen by many as an attack upon the gold OA movement [5.50]. Studies were rapidly undertaken analyzing the extent to which 'predatory journals' were affecting the whole gold OA publishing enterprise. One of the most significant findings was that most of their actions were geographically restricted to India and Nigeria [5.50, 51].

Moreover, South America is the continent with the largest share of gold OA publications (up to 74%) [5.52], but its presence in predatory journals represents barely 0.5% [5.51]. These findings demonstrate that gold OA publications are not necessarily of a lesser quality. Following what we observed in Figs. 5.1 and 5.2, we could hypothesize that national differences in gold OA impact are affected by disciplinary biases and the types of publishers of journals from these countries. *Ennas* and *Di Guardo* [5.49] already point out that "journals owned by UK and US publishers have a very strong and positive relation to the [positioning in the SCImago Journal Rank] ranking" and that "journals adopting a business model requiring a form of payment to publish tend to become top rated more than others". Similarly, *Laakso* and *Björk* [5.38] highlight the diversity of types of journal publishers, geographical regions and scientific disciplines.

Table 5.2 compares the average journal impact factor and category normalized citation impact (CNCI) of gold OA journals, non-gold OA journals and the total collection of journals. Interestingly, while there are no significant differences in the average journal impact factor, gold OA journals have on average a much lower category normalized citation impact than non-gold OA journals (0.62 against 0.85).

Figure 5.4a,b compare the citation impact of gold OA journals with that of other journals, broken down by discipline. They also indicate the absolute number of gold OA journals in a discipline. Figure 5.4a shows results based on the journal impact factor, and Fig. 5.4b on the category normalized citation impact. While Fig. 5.4a shows that in natural sciences and in humanities gold OA journals have on average higher impact factor values than other journals have, and lower values in the other disciplines, Fig. 5.4b reveals that CNCI values of gold OA sources are below those of other journals in *all* disciplines, the difference being largest for social sciences and humanities. Note that the lower number of OA journals in the humanities is largely due to the fact that WoS does not calculate the impact factor of journals indexed in the arts and humanities citation index.

Figure 5.5 further analyzes the CNCI of gold OA journals as a function of the country of the publisher of the journals. It reveals large differences in average CNCI between publishing countries. In the set of countries publishing more than 10 gold OA journals, those published from The Netherlands, Germany, USA and England tend to have high CNCI values. These countries host large international publishing houses. Typical examples of countries with relatively low CNCI values and at least five gold OA journals are Colombia, Mexico, Serbia, India and Brazil. This evidence further strengthens the suggestion that the relation between impact and gold OA could be related more to other factors such as type of publisher, country of the journal and the field of scope of the journal, rather than with the fact that they are OA journals. Indeed, this perception aligns well with the findings provided by *Chavarro* et al. [5.53] who showed that, in the case of Colombia, publishing in what they refer to as non-mainstream journals is not only common, but that these journals are different in purpose than other journals, as they tend to publish research findings in subjects related to local knowledge or as a means to bridge the international community and local communities.

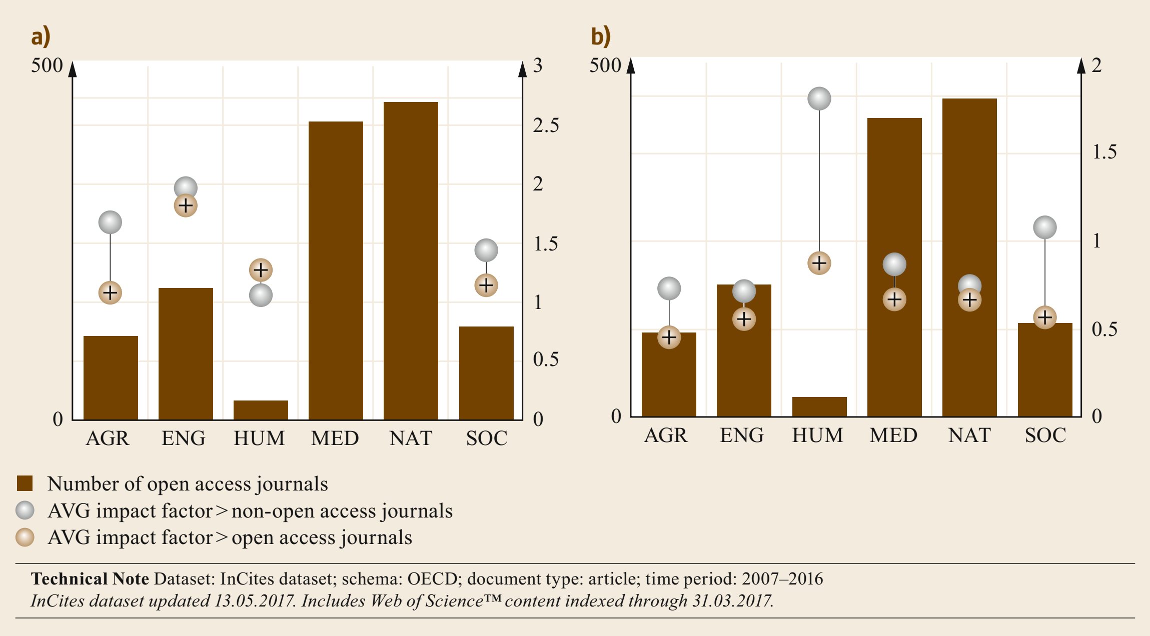


Fig. 5.4a,b Citation impact of gold OA journals with that of other journals, broken down by discipline. **(a)** Impact factor differences per area; **(b)** category normalized citation impact per area. AGR: agricultural science; ENG: engineering; HUM: humanities; MED: medical sciences; NAT: natural sciences; SOC: social sciences

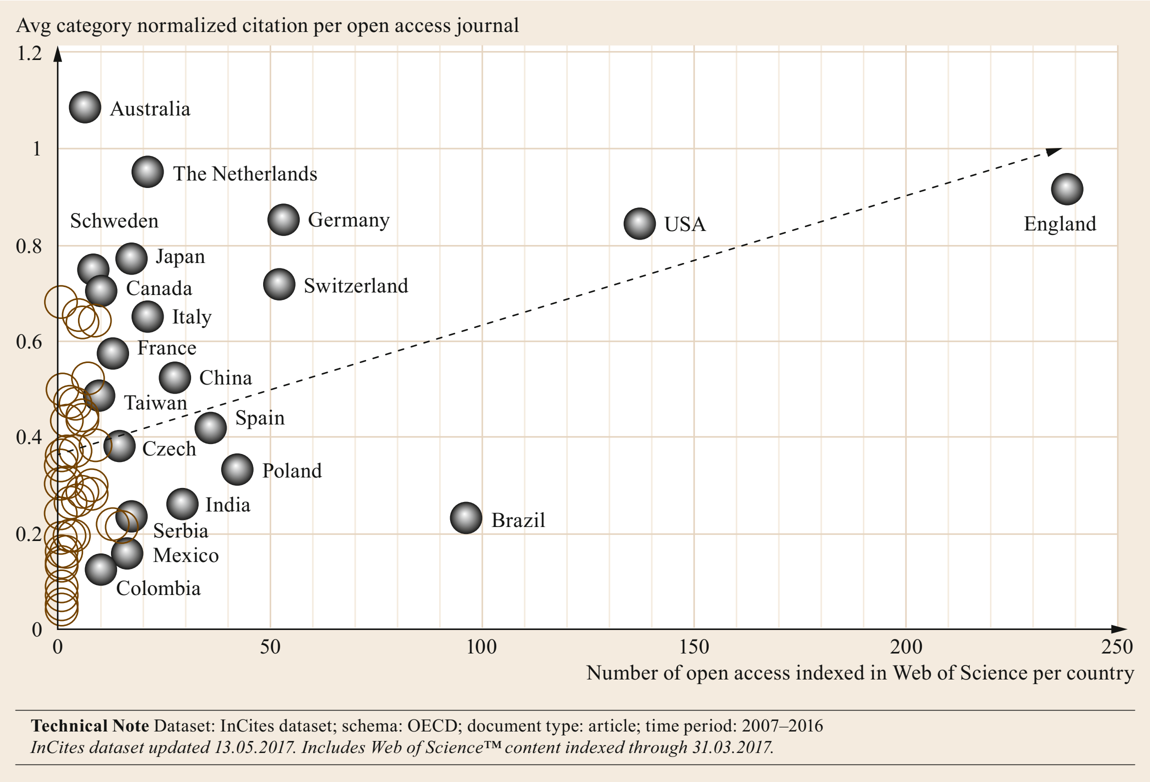


Fig. 5.5 Distribution and impact per country for OA journals indexed in the Web of Science

5.3.3 The Effect of OA Mega-Journals in the Publishing Ecosystem

It is impossible to discuss gold OA without taking some time to analyze the phenomenon of so-called OA mega-journals (OAMJs). Defined as [5.54]:

> Not only potentially disruptive in terms of altering the way research findings are assessed and communicated; [but] also disrupting academic culture itself

OAMJs have become, along with the rise of academic social media platforms [5.6], one of the major side-effects of gold OA in the publishing industry. These journals have not only impacted the scientific publishing culture due to the large number of papers they publish, but have also changed, or at least influenced, the ground rules of peer review by which a paper is considered to be worthy of publication.

According to *Wakeling* et al. [5.55], OAMJs are defined by two main characteristics:

1. They are broad in scope, covering many scientific areas
2. They set the peer review bar based on the technical soundness of manuscripts, not considering within their selection criteria the novelty of the publication or its significance and contribution to the field.

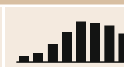
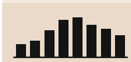
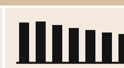
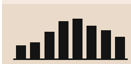
This has been seen by many as a lowering of the publication standards, and partially explains the perception by many researchers that these journals are of a lesser quality. What is more, big publishers have now introduced their own mega-journals (e. g., *Nature Communications*, *Scientific Reports* and *BMJ Open*). This move has been seen by some as a way "to tap into the stream of rejected manuscripts from their more selective top journals, in a system described as *cascading reviews*" [5.56]. Still, such malicious thinking does not seem supported by the data, which actually shows that, in terms of citations, both OAMJs and traditional journals show similar patterns [5.56].

Indeed, when examining the number of papers *PLOS One* publishes every year, the share they represent from the gold OA world output and its citation impact indicators, the numbers are quite revealing. Table 5.3 shows the publication trend of *PLOS One* for the 2009–2016 period. Starting with 4400 articles in the analyzed period, *PLOS One* reached its highest peak of articles published by year in 2013, when it produced more than 30 000 publications. This number was relatively stable in the two subsequent years with a decrease to 22 000 publications in 2016. *PLOS One* represented during the 2012–2014 period around 20% of all gold OA publications, decreasing in the later years. This was due to the decrease of publications, but also to the increase of publications from *Scientific Reports*, which went from over 3900 articles in 2014 to more than 10 700 articles in 2015 and up to 20 470 published papers in 2016.

A different issue is that related to the disruptive effect OAMJs have in scientific communities. According to *Beall* [5.27], a strong opponent of gold OA, "[t]hese journals, many of them now editorless, are losing the cohesion, soul, and community-binding roles that scholarly journals once had". Traditionally, scientific journals have been considered as niches which tie together and represent scientific communities, playing also a social role. OAMJs blur such communities, as contents of different areas and communities are published in the same place. However, where some

Part A | 5.3

Table 5.3 Articles published by *PLOS One* by year, contribution to gold OA overall publications and citation impact indicators. 2009–2016 period. Data retrieved from Web of Science

Publication year	2009	2010	2011	2012	2013	2014	2015	2016	TREND
Number of documents by *PLOS One* and contribution to the world									
WoS documents by *PLOS One*	4403	6728	13 780	23 441	31 492	30 038	28 114	22 077	
% Contribution of *PLOS One* to World's OA totals	7	9	15	20	22	18	15	12	
***PLOS One* impact factor and category normalized citation impact trend**									
Journal impact factor	4.35	4.41	4.09	3.73	3.53	3.23	3.06	–	
Category normalized citation impact	1.64	1.51	1.37	1.24	1.13	1.04	0.83	0.66	

Technical Note. Dataset: InCites dataset; schema: OECD; document type: article; time period: 2007–2016
InCites dataset updated 13.05.2017. Includes Web of Science™ content indexed through 31.03.2017

see a problem, others see an opportunity. *MacCallum* [5.57], acknowledges such a problem, but frames it from a different angle. OA means targeting not only scientists, but also policy makers, health managers, etc. OAMJs do not necessarily threaten the cohesion of scientific communities, but force publishers and information providers to rethink how to structure the increasing amount of literature produced by OAMJs, not only:

> To cater to different communities, but also to satisfy the needs of each individual and even enable them to generate new questions or discover novel avenues of research.

This is probably a somewhat opportunistic response to this issue, but one which reflects the disruptive effect OAMJs have had not only on the production of scientific literature, but also in its consumption.

5.4 Conclusions and Future Prospects

Despite the large number of studies devoted to defining, characterizing, analyzing and discussing OA, its integration into the scholarly communication system is still a grand challenge which leaves room for further debate and discussion. More than 25 years of OA have elapsed since the launch of ArXiv. Since then, topics under discussion have shifted many times. In the 1990s, studies tended to question copyright ownership [5.58] and explain the possibilities of making research findings more accessible [5.1]. The 2000s saw the expansion of OA, with many papers advocating for it [5.8, 13, 25, 59], explaining the different routes to OA [5.16] and debating about whether OA gave greater visibility to research literature [5.24, 41, 60, 61], in many cases arguing in favor of a citation advantage [5.21, 22, 42, 62, 63].

In the last decade, new topics have been added to this ongoing conversation. The settlement and growth of OA publishers, the emergence of OAMJs and of *predatory journals*, and the launch of academic social media platforms have led to more reflexive discussions as to the way OA is being integrated within the scientific communication system. Growing concerns as to how OA business models are affecting the quality of published research have been constantly present in the last few years [5.27, 28, 40]. This chapter intends to provide further insights as to the diversity of gold OA and its citation impact in comparison with non-gold OA. In this regard, the work of *Björk* and colleagues [5.15, 17, 38, 40, 44] already provides great in-depth analyses as to the heterogeneity of gold OA journals and the many factors that could be affecting the negative view OA journals seem to provoke within a large sector of the scientific community [5.26].

Building from their findings, we can relate the following comments. Regarding the overall number of gold OA publications and how much they represent from the overall number of publications, *Björk* et al. [5.15] indicated that "8.5% of all scholarly journal volume for 2008 is available through some form of Gold OA". The current study focused only on journal articles and it is based on data from Clarivate's InCites. We have obtained for 2008 a value of 4.37%, and for the year 2016 a value of 12.1%. This outcome illustrates that the overall share of gold OA output is still increasing with the emergence of new players, such as *Scientific Reports* and *Nature Communications*.

Björk and *Solomon* [5.40] concluded in their 2012 study that:

> OA journals indexed in Web of Science and/or Scopus are approaching the same scientific impact and quality as subscription journals, particularly in biomedicine and for journals funded by article processing charges.

In the current study, in which the category normalized citation impact (CNCI) of gold OA articles is compared to that of all other articles, a moderate increase is found in the ratio of the impact of gold OA and other types of articles. The question as to whether OA articles have a higher citation impact than non-OA papers has been addressed during the past 15 years in many studies analyzing the large multi-disciplinary citation indexes WoS and Scopus. *Laakso* et al. [5.17], *Björk* et al. [5.15], and *Björk* and *Solomon* [5.40] present reviews of these studies. To the best of the current authors' knowledge, the analysis presented in this chapter is one of the first to use the category normalized citation impact indicators at a large scale. The results show that gold OA articles have in the 2012–2016 period on average a citation impact that is some 15% lower than the world average impact in the gold OA articles' subject fields. It must be noted that this world average is based on all articles, both OA and non-OA.

The results presented in this chapter illustrate the heterogeneity in gold OA publishing, and the skewness of the underlying publication output and citation impact distributions. A limited number of gold OA journals accounts for a large percentage of the global gold OA output, and the effects of these journals upon overall scores can be assumed to be substantial.

These outcomes illustrate how cautious one should be in drawing generalized conclusions about gold OA publishing. The results obtained in the current study confirm the conclusions by *Björk* and *Solomon* [5.40] that:

> Gold OA publishing is rapidly increasing its share of the overall volume of peer-reviewed journal publishing, and there is no reason for authors not to choose to publish in OA journals just because of the 'OA' label, as long as they carefully check the quality standards of the journal they consider.

The relationship between access modality and citation impact is complex, and does not allow for simple, general conclusions. One should be cautious with generalizing statements such as *OA journals have higher (or lower) impact than subscription-based serials*. Based on the findings presented here, we have observed three models of gold OA production at the national level. These are shown in Table 5.4. The first model is that of countries like the USA, the United Kingdom, Germany and the Nordic countries, which publish in OA journals from big publishing firms with high impact factor, in many cases in OA mega-journals. The second model is exemplified in countries such as Brazil or India. These countries tend to have a large output in OA journals edited from their own countries. These journals tend to belong to specific fields (e.g., agricultural sciences in the case of Brazil), reinforcing the idea that they may be serving as bridging communities or focusing on topics of local or national interest [5.53]. The last model is represented by countries like Spain and Poland, which show a mixed combination of publishing in high impact OA mega-journals from big publishers, as well as publishing in nationally oriented OA journals from their own countries.

These results illustrate the many factors that could affect the final citation impact of OA publishing, and question statements against or in favor of OA publishing. Discussions related to gold OA are not independent of other factors such as the disciplinary profile of countries' output, national characteristics or types of publishers.

Table 5.4 Examples of countries representing three models of gold OA publishing

Journal name	Publisher country	Articles	JIF	CNCI
Model 1 – UK – Publication in English language and high impact factor OA journals				
PLOS One	USA	12 948	3.057	1.14
Scientific Reports	England	3356	5.228	1.28
Nature Communications	England	2131	11.329	3.48
BMJ Open	England	2020	2.562	0.82
Journal of High Energy Physics	Italy	1161	6.023	2.19
BMC Public Health	England	972	2.209	0.87
Nucleic Acids Research	England	945	9.202	6.36
Frontiers in Psychology	Switzerland	863	2.463	1.2
Trials	England	757	1.859	0.76
Atmospheric Chemistry and Physics	Germany	686	5.114	2.24
PLOS Neglected Tropical Diseases	USA	679	3.948	1.8
PLOS Genetics	USA	601	6.661	2.56
New Journal of Physics	England	595	3.57	1.34
Malaria Journal	USA	580	3.079	1.33
PLOS Pathogens	USA	543	7.003	2.58

Table 5.4 (continued)

Journal name	Publisher country	Articles	JIF	CNCI
MODEL 2 – Brazil – Publication in national and low impact factor OA journals				
PLOS One	USA	3961	3.057	0.86
Semina-Ciencias Agrarias	Brazil	1783	0.229	0.5
Ciencia Rural	Brazil	1746	0.376	0.22
Ciencia and Saude Coletiva	Brazil	1501	0.669	0.23
Arquivo Brasileiro de Medicina Veterinaria e Zootecnia	Brazil	1192	0.21	0.16
Pesquisa Veterinaria Brasileira	Brazil	1057	0.335	0.23
Quimica Nova	Brazil	1013	0.617	0.16
Cadernos de Saude Publica	Brazil	950	0.92	0.3
Journal of the Brazilian Chemical Society	Brazil	926	1.096	0.32
Revista Brasileria de Engenharia Agricola e Ambiental	Brazil	889	0.478	0.25
Pesquisa Agropecuaria Brasileira	Brazil	871	0.564	0.29
Revista Brasileira de Ciencia do Solo	Brazil	748	0.611	0.36
Revista da Escola de Enfermagem da USP	Brazil	748	0.415	0.15
Anais da Academia Brasileira de Ciencias	Brazil	721	0.717	0.27
Brazilian Journal of Biology	Brazil	686	0.559	0.18
Journal name	**Publisher country**	**Articles**	**JIF**	**CNCI**
MODEL 3 – Spain – Publication in both, national and English language OA journals				
PLOS One	USA	5473	3.057	1.07
Scientific Reports	England	1273	5.228	1.37
Nutricion Hospitalaria	Spain	892	1.497	0.3
Sensors	Switzerland	868	2.033	0.7
Journal of High Energy Physics	Italy	613	6.023	2.45
Nature Communications	England	608	11.329	3.47
Anales de Psicologia	Spain	490	0.574	0.36
Gaceta Sanitaria	Spain	376	1.509	0.36
Physics Letters B	The Netherlands	362	4.787	3.48
Psicothema	Spain	340	1.245	0.62
New Journal of Physics	England	333	3.57	1.38
Spanish Journal of Agricultural Research	Spain	322	0.76	0.37
Nucleic Acids Research	England	317	9.202	1.9
Informes de la Construccion	Spain	281	0.227	0.12
BMC Genomics	England	276	3.867	1.28
Nefrologia	Spain	276	1.207	0.35

References

5.1 P. Ginsparg: Winners and losers in the global research village, Ser. Libr. **30**(3/4), 83–95 (1997)

5.2 M. Kurtz, T. Brody: The impact loss to authors and research. In: *Open Access: Key Strategic, Technical and Economic Aspects*, ed. by N. Jacobs (Chandos, Oxford 2006)

5.3 J. Beall: Predatory publishers are corrupting open access, Nature **489**(7415), 179 (2012)

5.4 J. Esposito: The devil you don't know: The unexpected future of open access publishing, First Monday (2004), https://doi.org/10.5210/fm.v9i8.1163

5.5 F.J. Friend: Big deal–good deal? Or is there a better deal?, Learn. Publ. **16**(2), 153–155 (2003)

5.6 B.-C. Björk: The open access movement at a crossroad: Are the big publishers and academic social media taking over?, Learn. Publ. **29**(2), 131–134 (2016)

5.7 J. Whitfield: Elsevier boycott gathers pace, Nat. News (2012), https://doi.org/10.1038/nature.2012.10010

5.8 A. Swartz: Guerilla open access manifesto, http://wavelets.ens.fr/BOYCOTT_ELSEVIER/DECLARATIONS/DECLARATIONS/2008_07_01_Aaron_Swartz_Open_Access_Manifesto.pdf (2008)

5.9 D.S. Chawla: Unpaywall finds free versions of paywalled papers, Nat. News (2017), https://doi.org/10.1038/nature.2017.21765

5.10 J. Bohannon: Who's downloading pirated papers? Everyone, Science **352**(6285), 508–512 (2016)

5.11 P. Suber: Removing the barriers to research: An introduction to open access for librarians, College Res. Libr. News **64**, 92–94 (2003)

5.12 P. Suber: Open access, impact, and demand: Why some authors self archive their articles, BMJ **330**(7500), 1097 (2005)
5.13 A. Zuccala: The lay person and open access, Annu. Rev. Inf. Sci. Technol. **43**(1), 1–62 (2009)
5.14 B.-C. Björk: Open access subject repositories: An overview, J. Assoc. Inf. Sci. Technol. **65**(4), 698–706 (2014)
5.15 B.-C. Björk, P. Welling, M. Laakso, P. Majlender, T. Hedlund, G. Guðnason: Open access to the scientific journal literature: Situation 2009, PLOS ONE **5**(6), e11273 (2010)
5.16 S. Harnad, T. Brody, F. Vallières, L. Carr, S. Hitchcock, Y. Gingras, E.R. Hilf: The access/impact problem and the green and gold roads to open access: An update, Ser. Rev. **34**(1), 36–40 (2008)
5.17 M. Laakso, P. Welling, H. Bukvova, L. Nyman, B.-C. Björk, T. Hedlund: The development of open access journal publishing from 1993 to 2009, PLOS ONE **6**(6), e20961 (2011)
5.18 É. Archambault, D. Amyot, P. Deschamps, A. Nicol, F. Provencher, L. Rebout, G. Roberge: Proportion of open access papers published in peer-reviewed journals at the European and world levels—1996–2013, http://digitalcommons.unl.edu/scholcom/8/ (2014)
5.19 J.P. Tennant, F. Waldner, D.C. Jacques, P. Masuzzo, L.B. Collister, C.H.J. Hartgerink: The academic, economic and societal impacts of open access: An evidence-based review, F1000Research **5**, 632 (2016)
5.20 R. Van Noorden: Open access: The true cost of science publishing, Nat. News **495**(7442), 426 (2013)
5.21 G. Eysenbach: Citation advantage of open access articles, PLOS Biology **4**(5), e157 (2006)
5.22 Y. Gargouri, C. Hajjem, V. Larivière, Y. Gingras, L. Carr, T. Brody, S. Harnad: Self-selected or mandated, open access increases citation impact for higher quality research, PLOS ONE **5**(10), e13636 (2010)
5.23 S. Harnad, T. Brody: Comparing the impact of open access (OA) vs. non-OA articles in the same journals, D-Lib Mag (2004), https://doi.org/10.1045/june2004-harnad
5.24 H.F. Moed: The effect of "open access" on citation impact: An analysis of ArXiv's condensed matter section, J. Am. Soc. Inf. Sci. Technol. **58**(13), 2047–2054 (2007)
5.25 W. Crawford: Free electronic refereed journals: Getting past the arc of enthusiasm, Learn. Publ. **15**(2), 117–123 (2002)
5.26 A.A. Agrawal: Four more reasons to be skeptical of open-access publishing, Trends Plant Sci. **19**(3), 133 (2014)
5.27 J. Beall: The open-access movement is not really about open access, TripleC Commun. Cap. Crit. Open Access J. Glob. Sustain. Inf. Soc. **11**(2), 589–597 (2013)
5.28 J. Bohannon: Who's afraid of peer review?, Science **342**(6154), 60–65 (2013)
5.29 T.N. van Leeuwen, C. Tatum, P. Wouters: Open access publishing and citation impact–An international study. ISSI, https://pdfs.semanticscholar.org/55f9/1f879fe7867da70b782914ac78d6fc7c6984.pdf (2015)
5.30 R. Van Noorden: Half of 2011 papers now free to read, Nat. News **500**(7463), 386 (2013)
5.31 R. Cullen, B. Chawner: Institutional repositories, open access, and scholarly communication: A study of conflicting paradigms, J. Acad. Librariansh. **37**(6), 460–470 (2011)
5.32 P. Suber: Gratis and libre open access, https://dash.harvard.edu/handle/1/4322580 (2008)
5.33 B.-C. Björk: Gold, green, and black open access, Learn. Publ. **30**(2), 173–175 (2017)
5.34 H. Piwowar, J. Priem, V. Larivière, J.P. Alperin, L. Matthias, B. Norlander, A. Farley, J. West, S. Haustein: The state of OA: A large-scale analysis of the prevalence and impact of open access articles, PeerJ **6**, e4375 (2018)
5.35 Y. Gargouri, V. Lariviere, Y. Gingras, L. Carr, S. Harnad: Green and gold open access percentages and growth, by discipline. In: *17th Int. Conf. Sci. Technol. Indic. (STI)* (2012)
5.36 C. Gumpenberger, M.-A. Ovalle-Perandones, J. Gorraiz: On the impact of gold open access journals, Scientometrics **96**(1), 221–238 (2013)
5.37 M. Jubb, J. Cook, D. Hulls, D. Jones, M. Ware: Costs, risks and benefits in improving access to journal articles, Learn. Publ. **24**(4), 247–260 (2011)
5.38 M. Laakso, B.-C. Björk: Anatomy of open access publishing: A study of longitudinal development and internal structure, BMC Medicine **10**, 124 (2012)
5.39 A. Cochran: Interview with Thomson Reuters: InCites platform offers new analytics and transparency, https://scholarlykitchen.sspnet.org/2014/08/08/qa-with-thomson-reuters-incites-platform-offers-new-analytics-and-transparency/ (2014)
5.40 B.-C. Björk, D. Solomon: Open access versus subscription journals: A comparison of scientific impact, BMC Medicine **10**, 73 (2012)
5.41 T. Brody, S. Harnad, L. Carr: Earlier web usage statistics as predictors of later citation impact, J. Am. Soc. Inf. Sci. Technol. **57**(8), 1060–1072 (2006)
5.42 I.D. Craig, A.M. Plume, M.E. McVeigh, J. Pringle, M. Amin: Do open access articles have greater citation impact?: A critical review of the literature, J. Informetrics **1**(3), 239–248 (2007)
5.43 A. Swan: The Open Access Citation Advantage: Studies and Results to Date, http://eprints.soton.ac.uk/268516 (2010)
5.44 D.J. Solomon, M. Laakso, B.-C. Björk: A longitudinal comparison of citation rates and growth among open access journals, J. Informetrics **7**(3), 642–650 (2013)
5.45 R. Meneghini, R. Mugnaini, A.L. Packer: International versus national oriented Brazilian scientific journals, a scientometric analysis based on SciELO and JCR-ISI databases, Scientometrics **69**(3), 529–538 (2006)
5.46 D. Torres-Salinas, N. Robinson-Garcia, I.F. Aguillo: Bibliometric and benchmark analysis of gold open access in spain: Big output and little impact, Informetria **25**(1), 17–24 (2016)
5.47 M.E. McVeigh: *Open Access Journals in the ISI Citation Databases: Analysis of Impact Factors and Citation*

Part A | 5

Patterns: A Citation Study from Thomson Scientific (Thomson Scientific, Philadelphia 2004)
5.48 E. Giglia: The impact factor of open access journals: Data and trends, http://eprints.rclis.org/14666 (2010)
5.49 G. Ennas, M.C. Di Guardo: Features of top-rated gold open access journals: An analysis of the scopus database, J. Informetrics **9**(1), 79–89 (2015)
5.50 J. Xia, J.L. Harmon, K.G. Connolly, R.M. Donnelly, M.R. Anderson, H.A. Howard: Who publishes in "predatory" journals?, J. Assoc. Inf. Sci. Technol. **66**(7), 1406–1417 (2015)
5.51 C. Shen, B.-C. Björk: 'Predatory' open access: A longitudinal study of article volumes and market characteristics, BMC Medicine **13**, 230 (2015)
5.52 S. Miguel, Z. Chinchilla-Rodriguez, F. de Moya-Anegón: Open access and scopus: A new approach to scientific visibility from the standpoint of access, J. Am. Soc. Inf. Sci. Technol. **62**(6), 1130–1145 (2011)
5.53 D.A. Chavarro, P. Tang, I. Rafols: Why researchers publish in non-mainstream journals: Training, knowledge bridging, and gap filling, Res. Policy **46**(9), 1666 (2017)
5.54 V. Spezi, S. Wakeling, S. Pinfield, C. Creaser, J. Fry, P. Willett: Open-access mega-journals: The future of scholarly communication or academic dumping ground? A review, J. Documentation **73**(2), 263–283 (2017)
5.55 S. Wakeling, P. Willett, C. Creaser, J. Fry, S. Pinfield, V. Spezi: Open-access mega-journals: A bibliometric profile, PLOS ONE **11**(11), e0165359 (2016)
5.56 B.-C. Björk, P. Catani: Peer review in megajournals compared with traditional scholarly journals: Does it make a difference?, Learn. Publ. **29**(1), 9–12 (2016)
5.57 C.J. MacCallum: Why one is more than 5, PLOS Biology **9**(12), e1001235 (2011)
5.58 S. Bachrach, R.S. Berry, M. Blume, T. Foerster, A. von Fowler, P. Ginsparg, A. Moffat: Who should own scientific papers?, Science **281**(5382), 1459–1460 (1998)
5.59 P. Ginsparg: As we may read, J. Neuroscience **26**(38), 9606–9608 (2006)
5.60 T.F. Frandsen: The effects of open access on unpublished documents: A case study of economics working papers, J. Informetrics **3**(2), 124–133 (2009)
5.61 A. Haque, P. Ginsparg: Positional effects on citation and readership in arXiv, J. Am. Soc. Inf. Sci. Technol. **60**(11), 2203–2218 (2009)
5.62 P.M. Davis, M.J. Fromerth: Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles?, Scientometrics **71**(2), 203–215 (2007)
5.63 P.M. Davis, B.V. Lewenstein, D.H. Simon, J.G. Booth, M.J.L. Connolly: Open access publishing, article downloads, and citations: Randomised controlled trial, BMJ **337**, a568 (2008)

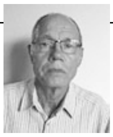
Daniel Torres-Salinas
EC3metrics SL y Medialab-UGR
Universidad de Granada
Granada, Spain
torressalinas@ugr.es

Daniel Torres-Salinas holds a PhD in scientific documentation. He currently works as a bibliometrician at the Universities of Granada and Navarra. He is a researcher in the EC3 Research Group (Evaluación de la Ciencia y de la Comunicación Científica) and focuses on bibliometric topics. In addition, Daniel Torres-Salinas is CEO of the EC3metrics spin-off. He is co-author of various evaluation tools and has published more than 50 scientific.

Nicolas Robinson-García
School of Public Policy
Georgia Institute of Technology
Atlanta, GA, USA
elrobinster@gmail.com

Nicolas Robinson-García is a Postdoctoral Researcher at the School of Public Policy of the Georgia Institute of Technology in the United States. He holds a PhD in Social Sciences from the University of Granada (Spain). He is currently interested in the development of quantitative methods to analyze scientific mobility flows as well as the analysis of new data sources and methods to trace societal engagement of researchers.

Henk F. Moed
Amsterdam, The Netherlands
hf.moed@gmail.com



Henk F. Moed obtained a PhD in Science Studies at the University of Leiden in 1989 and was a senior staff member at the Centre for Science and Technology Studies (CWTS) at Leiden University between 1981 and 2010. From 2010 to 2014 he was Senior Scientific Advisor at the international publishing company Elsevier, Amsterdam, The Netherlands. He has been active in numerous research topics in the fields of bibliometric research and assessment of research performance
Henk Moed has published over 100 research articles in international, peer-reviewed journals, has been program chair of numerous international conferences in the field, and is a member of the Editorial Boards of several journals in his field. He is currently an independent scientific advisor and Visiting Professor at the Sapienza University of Rome and the University of Granada.

6. Science Forecasts: Modeling and Communicating Developments in Science, Technology, and Innovation

Katy Börner, Staša Milojević

In a knowledge-based economy, science and technology are omnipresent, and their importance is undisputed. Equally evident is the need to allocate resources, both monetary and human, in an effective way to foster innovation [6.1, 2]. In the preceding decades, science policy has embraced data mining and metrics to gain insights into the structure and evolution of science and to devise metrics and indicators [6.3], but it has not invested significant efforts into mathematical, statistical, and computational models that can predict future developments in science, technology, and innovation (STI) in support of data-driven decision making.

Recent advances in computational power combined with the unprecedented volume and variety of data concerning science and technology developments (e.g., publications, patents, funding, clinical trials, and stock market and social media data) yielded ideal conditions for the advancement of computational modeling approaches that can be not only empirically validated, but used to simulate and understand the structure and dynamics of STI in support of improved human decision making.

In this chapter, we review and demonstrate the power of computational models for simulating and predicting possible STI developments and futures. In addition, we discuss novel means to visualize and broadcast STI forecasts to make them more accessible to general audiences.

6.1 Models and Visualizations

Science, technology, and innovation are crucial for the prosperity of nations, and are a driving force of human civilization. After World War II, science entered a phase of accelerated growth, reflected in the exponential rise in the number of active scientists and an increase of scientific output [6.4]. Science itself is undergoing a transformation, with most researchers engaging in collaborative or team work [6.5, 6]. In order to create effective science policies and maximize the returns on our society's investments in STI, we must understand STI as a complex and dynamic system that emerges from interdependences and interactions of different actors at different levels of aggregation.

Models of STI aim to inform policy decision making in many fields, including education, energy, healthcare, security, and others [6.7, 8]. These models do not replace—but rather empower—experts to make better informed decisions when selecting reviewers, picking the best proposals for funding, or when making resource allocation decisions. They are a new kind of

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_6

'macroscope tool' [6.9] that help derive key insights from big data in support of evidence-based policy.

Some existing models of STI are optimized to make recommendations. IBM's *Watson*, for example, can suggest reviewers for a set of proposals without much information on the type of match or the matching process. Other models aim to capture the true structure and dynamics of complex STI systems, simulating the diffusion of ideas and experts, estimating the impact of population explosion and aging, or communicating the probable outcomes of different policy decisions. Still others help answer either resource allocation or multifaceted strategic questions, the latter of which are often used in a team setting where small multidisciplinary groups investigate and debate alternative futures together.

Computational models are well established in many fields: meteorology, where they are used to predict weather and storms; epidemiology, to predict and prevent pandemics; and climate, to predict future scenarios and set carbon prices. In industry (hereafter used as a general term to indicate the various industrial sectors, such as retail, IT, car manufacturing, etc.), computational models are used to optimize operations, management, production, distribution, and marketing. Early adopters of data-driven decision making (most notably Target, Walmart, and Amazon) now dominate their sectors. Those who were slow to invest and then did so in isolated aspects of the organization (most notably Sears and Kmart) are headed towards bankruptcy.

Interactive data visualizations that show probable futures in response to different policy decisions or external events can help stakeholders discuss model assumptions, designs, and outputs. Ideally, stakeholders get to "drive the future before it is implemented" [6.10, 11]; they can quickly explore different policy options and discard those that lead to undesired consequences [6.2, 12]. However, designing effective interfaces that let different stakeholders communicate and explore different scenarios is a nontrivial endeavor.

6.2 Models and Modeling

While our world is infinitely complex, our ability to sense, understand, and act within that world is finite. To capture and interpret the structures and dynamics of a complex system such as STI, scientists build models, which are simplified representations of a system [6.13]. Models bring conceptual unity to what is otherwise too complex to understand and manage. Regardless of the approach, the goal of any model is to simplify thinking "while still retaining some ability to illuminate reality" [6.14, p. 11]. In order to understand and predict different aspects of the world, models reduce the world to a subset of elements and laws that govern the behavior of those elements. Such simplification allows researchers to focus on and elucidate only the specific elements of a system that concern them. While every model is bounded by its initial framework, this does not mean that it cannot increase our understanding of the phenomenon at hand.

Complex systems research, however, challenges the notion that by perfectly understanding the behavior of each component of a system, we will understand the system as a whole. While there is no agreed upon definition of complex systems, the combination of various definitions leads to the following characteristics that a system needs to have in order to be considered complex [6.15, 16]. Two major components of a system are its entities and the interactions among those entities, with a much heavier emphasis on the interactions than on the entities.

A complex system usually has a large number of entities that mainly respond to local information (i. e., each element of the system is ignorant of the behavior of the system as a whole). The interactions these entities have can be: nonlinear (small changes in system variables that can have disproportionate outcomes); dynamic (changes over time); rife with feedback loops (both positive and negative, which can lead to distributions such as a power law); fairly rich (any element in the system influences and is influenced by quite a few others); and fairly short range. So far as the system as a whole is concerned, it is open (i. e., interacts with its environment); requires nonequilibrium conditions (there needs to be a constant flow of energy to ensure the survival of the system); and has a history (not only does it evolve through time, but its past is coresponsible for the present behavior).

Computational models consist of input (theories translated first into mathematical equations, and then into algorithms with different parameter values) and output (structures, or the behavior of the model over time). A dynamic system is one which by its very nature changes its state in ways that can be modeled by an application of an evolution law (a set of rules that describe what phase space configuration the system will occupy in the next moment). In the case of complex systems, these rules—though often very simple—can lead to so-called *emergent behavior*, a phenomenon in which "individual, localized behavior aggregates into global behavior" [6.14, p. 44]. When modeling the evolution of dynamic complex systems, researchers must remember that the resultant model represents one configuration in a phase space, given at time t [6.17], and as the models aim to capture system dynamics, output at time t often serves as input for computing time $t+1$.

Typically, the accuracy of simulations increases with both the ease of repetition and the number of simulations run (i. e., the number of possible futures obtained). Running simulations multiple times allows for better estimates concerning the sensitivity of outcomes to initial conditions, as well as the probabilities associated with those outcomes.

Computer modeling is gaining traction as an acceptable approach to doing science in a wide range of fields, from astronomy to economics [6.18]. Computational models use simulations to study the behavior of a system, and these simulations pose new questions regarding the scientific method, the nature of evidence, theory and theory building, and the role of data [6.19]. Outside pressures have forced researchers in climatology—a field heavily dependent on computer models—to be at the forefront of deeply critical thinking regarding the capabilities and limitations of computer modeling [6.20]. In this way, climatology exemplifies how community-lead, large-scale endeavors to gather, model, and visualize data can lead to significant infrastructure building.

Using simulations in the social sciences is a more recent phenomenon; however, a number of excellent resources showcase the benefits of broad usage cases [6.21] and provide practical guidance [6.22, 23]. Computational models can be used both to advance theory via conceptual models, and as tools to enhance decision making via predictions. In both cases, modeling is an iterative process that includes both induction and deduction [6.14] to revise and improve an initial set of assumptions, often leading to better results.

Recent developments in machine learning have dramatically improved researchers' capabilities to identify structures and patterns in data to aid with decision making [6.24, 25]. *Jon Kleinberg* and colleagues [6.26] provide a great overview of what they call "prediction policy problems" ranging from the medical field (predicting which surgeries will be futile) to criminal justice (deciding on whether to detain or release an arrestee).

6.3 Modeling Science

The book *Models of Science Dynamics* [6.8] provides a unique review of major model classes—from population dynamics models to complex network models—accessible to science policy researchers and practitioners. Two special issues in *Scientometrics* entitled *Modeling Science: Studying the Structure and Dynamics of Science* [6.27] and *Simulating the Processes of Science, Technology, and Innovation* [6.28] feature research, applications, and validations of exemplary STI models.

Models capturing the structure and evolution of scientific endeavor fall into one of two categories: descriptive and predictive [6.29]. Descriptive models include maps, and aim to describe the major features of static datasets. Predictive (or process) models aim to capture the mechanisms and temporal dynamics by which real-world systems evolve, focusing on the identification of elementary mechanisms that lead to the emergence of specific structures or dynamics. Ultimately, process models seek to simulate, statistically describe, or formally reproduce statistical characteristics of interest.

Computational models have been developed to enhance our knowledge of fundamental generating processes regarding citing, publishing, careers, rewards, funding, team formation, problem selection, and research areas dynamics. They gained traction in recent years because of their power to simulate processes leading to particular outcomes. Particularly important were findings that identified universal patterns, that is, patterns holding across majority of scientific fields, such as citations dynamics and timing of major discoveries.

A number of scientific models draw from complexity theory. Modern notions of complexity have their roots in theories of chaos, complex systems, fractal geometry, nonlinear dynamics, and self-organizing criticality. Complexity theory has been influential in physical science, technology, and mathematics for some time. This influence is newer and less developed in social science [6.30]. This is unsurprising, since models of complex systems work best with homogeneous elements where there is little or no difference between the individual elements of the system (e. g., atoms and molecules). Using the tools of complexity theory to cover the richness of both entities and their relationships within social systems proves to be a harder task.

The *natural* and *self-organizing* development of science towards more interdisciplinary activities is comparable with ecological systems that exhibit growth and emergent behavior [6.31, 32]. *Anthony van Raan* [6.33] expanded on this idea, portraying science not only as an interdisciplinary, complex, and self-organizing system, but as an amalgam of "cognitive regions" derived from the parts of pre-established disciplines. Such disciplines represent research fields that originated from earlier interdisciplinary developments.

In this model, science itself is a living, complex and dynamic system consisting of several ever-growing subsystems, each of which unfolds further into a myriad of different fields and subfields. A variation of the so-called "epidemics model" [6.34] was used in the 1960s to develop an epidemic theory on the diffusion of ideas and the growth of scientific specialties [6.35]. By using mast cell research as a case study, William Goffman demonstrated that it was possible to see growth and development as sequences of overlapping epidemics.

A wide range of studies used network-based models of citations to understand different aspects of science. A number of studies focused on understanding the dynamics of citation accumulation, starting from the identification of cumulative advantage/preferential attachment as the driving mechanism behind the power-law distribution of citations [6.36, 37]. *Filippo Radicchi* et al. [6.38] found that citation distributions are universal across fields by replacing the raw number of citations with relative ones. More advanced models of citation dynamics have focused on features such as the obsolescence of knowledge, which leads to a decrease in the number of citations as a function of time [6.39, 40].

Dashun Wang et al. [6.41] have developed this idea the furthest, developing a generative model that takes into account three parameters (the number of previous citations, obsolescence, and fitness) to predict citation dynamics of individual papers. Such network approaches are also used to identify communities of research papers that frequently cite one another [6.42], a task of enormous importance for policy and evaluation. These citation networks were also used to trace the usage of words and phrases to determine whether the usage of such words corresponds to the emergence of new paradigms [6.43].

While most network models focus on a single type of node at a time, some combine a number of different types of nodes. For example, work by *Feng Shi* et al. [6.44] aims to understand how choices at the microlevel (e. g., an individual scientist's choice of topics) may constrain the development and advancement of knowledge at the macrolevel, making incremental advances/improvements to the things that are already known, rather than huge leaps to unconnected—and therefore still unimagined—futures.

A number of models focus on scientific careers. For example, *Alexander Petersen* et al. [6.45] developed a generative model showing the detrimental effect policy decisions related to the increased availability of short-term positions have on researchers' productivity levels. Work by *Albert-László Barabási* and his team uses a stochastic model to show that the timing of one's most important contribution is not the result of (academic) age, but can occur at any stage of one's career, and is a function of productivity [6.46].

Agent-based models can reveal the microprocesses of individuals that lead to particular macrolevel patterns. These models focus on the relations and interactions among entities rather than the characteristics of the entities themselves. *Nicholas Payette* [6.47] provides an excellent overview of agent-based models in the context of studying science. *Nigel Gilbert* introduced the first agent-based model to study science focused on papers (rather than authors) as agents and managed to reproduce Lotka's law of productivity, as well as the rise and decline of specializations [6.48].

Katy Börner et al. [6.49] developed a more nuanced model called TARL (topics, aging, and recursive linking), in which they simulate the simultaneous coevolution of networks of papers and authors driven by the "rich-get-richer" phenomenon [6.37, 50, 51]. *Xiaoling Sun* et al. [6.52] revisited the topic of the growth and decline of disciplines, proposing an agent-based model in which the evolution of disciplines is guided mainly by social interactions of scientists writing papers together. The disciplines thus rise and fall through the splitting and merging of communities of collaborators.

A key insight to be drawn from existing model results is that science is complex, and therefore the study of science is also complex. This complexity comes from the fact that not only are communication processes under study multilayered, but that both the data and the latent structures within the data are evolving over time. Furthermore, it is now obvious that the intricate relationships between social, conceptual, cognitive, and institutional forces need to be taken into account to fully understand the organization of science.

Despite great advances towards expanding our understanding of science, there are definite limitations in terms of predicting the emergence of a new field [6.53]. These deficiencies are due to the fact that 'normal science' is much easier to predict than 'radical innovations'. External forces (new policies, wars, etc.) have a major impact on the development of science, while data access and model development are both limited.

Current work focuses on the expansion of data sources from the traditional output of research in the forms of bibliographic data on publications, grants, and patents to the analysis of full text of those resources, grant applications (both successful and unsuccessful), mentorship (formal and informal), conferences, employment data, social media, etc. [6.53]. In parallel, algorithm development aims to capture not only strong associations but also causation [6.53, 54]. One step in that direction is using counterfactual scenarios to assess how well models perform [6.26, 55–57].

6.4 Exemplary Models of Science

This section discusses two science models in more detail. The first example describes how teams in various fields have evolved over time, and what it is they contribute to contemporary science. The second example proposes radical changes to the current funding system. Both of these models have been empirically validated, and reveal a high correlation between the simulated datasets and the structures/dynamics found in publication and funding data.

6.4.1 The Importance of Small Teams in the Big Science Era

Contemporary science is a collaborative effort within an intricate network of people, institutions, concepts, and technology. Many projects are of such complexity and scope that they require the joint efforts of many individuals with diverse expertise, culminating in teams of hundreds. Furthermore, studies suggest that large, interdisciplinary teams are more likely to produce high-impact work.

Yet only 50 years ago, the situation was very different. Most papers were written by single authors, and the largest coauthor teams did not exceed ten members. How did this change in the production of knowledge occur? How do science teams form, and what processes lead to their expansion? Perhaps most importantly, what makes a successful team?

Considering these questions, team size distribution lies at the heart of our understanding of collaborative practices and research productivity. As Fig. 6.1 shows, knowledge production today is qualitatively different from that of earlier times: *little science* performed by individuals or small groups of researchers is largely superseded by *big science* efforts conducted by large teams that span disciplinary, institutional, and national boundaries.

In Fig. 6.1, we see a change in the distribution of research team sizes in physics from a Poisson distribution to one dominated by a fat tail (a power law). In 1941–1945, for each paper with five authors, there were one thousand single-authored papers (blue). In 2006–2009, there were as many papers with five authors as there were single authored papers (red), and very large teams were not uncommon. Such a distribution (Fig. 6.1) can be reproduced using the model developed by *Staša Milojević* [6.5], which demonstrated how teams emerge, grow, and would evolve in the future.

Vitally, Milojević's model shows that team formation was, and remains, a Poisson process resulting in relatively small core teams (including single-investigator teams) carrying out certain types of research. The model also simulates the emergence of larger teams over the last 50 years in all fields of science, albeit with varying pace and magnitude of change.

According to the Milojević model, every big team originates from a small team; while some small teams do not change in size, others quickly accumulate additional members proportionally to the past productivity

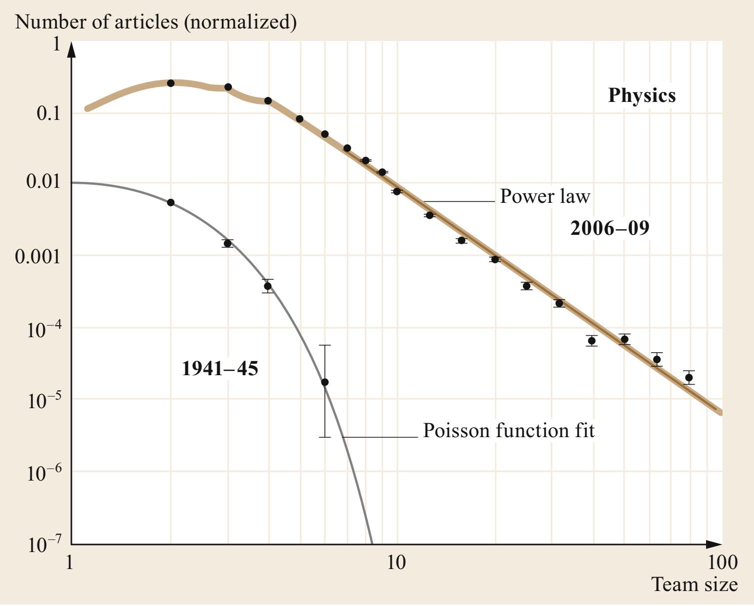


Fig. 6.1 Change in the distribution of team sizes

of preexisting team members, eventually allowing small teams to grow into big teams.

Furthermore, Milojević's model shows that relatively small teams dominate knowledge production in most fields; cumulatively, small teams still contribute more new knowledge than large teams. These findings are of key importance to policy, because they show that increased funding emphasis on large teams may undermine the very process by which large, successful teams emerge.

6.4.2 Crowdsourcing Funding Allocation

As funding agencies consume resources that could be more productively used to conduct and finance research, *Johan Bollen* et al. [6.58] argue that scholars "invest an extraordinary amount of time, energy and effort into the writing and reviewing of research proposals". In their 2014 paper, they used National Science Foundation (NSF) and Taulbee Survey data to calculate the return on investment for scholars in computer science. This calculation reveals a negative return on investment.

Given a computer and information science and engineering (CISE) funding rate of 21%, four professors working full-time for four weeks on a proposal submission with labor costs of about \$35 000, five submission-review cycles may be required, resulting in a total expected labor cost of \$175 000.

The average NSF grant is \$164 526 per year, to which US universities charge about 50% of their overhead, leaving roughly \$109 684 for research. Consequently, the four professors in question lose \$65 316 of paid research time by obtaining a grant. US universities might even forbid professors to apply for grants—if they can afford to forgo the indirect dollars. Note that this simple calculation does not cover any time spent by scholars to review proposals. In 2015 alone, NSF conducted 231 000 proposal reviews to evaluate 49 600 proposals.

Bollen et al. [6.58] then go on to propose a *FundRank* model to (partially) replace the current process of government research funding allocation by expert-based crowdsourcing. In this new FundRank system, each eligible scholar (e. g., all eligible to submit NSF and National Institutes of Health (NIH) grants today) receives a certain dollar amount each year—let's say \$100 000. She then needs to give a certain fraction (e. g., 50%) to colleagues that are most deserving by logging into a centralized website and entering names and amounts. In this way, scholars collectively assess each other's merit and *fund-rank* one another, with high ranking scholars receiving the most funding.

Instead of spending weeks writing and reviewing proposals, scholars are now incentivized to spend time communicating the value and impact of their past, current, and planned work so that others can judge their contributions. Using a fully digital system, conflicts of interest could be easily identified and honored; networks of mutual favors could be detected automatically, and results shared publicly.

FundRank was implemented using the recursive PageRank algorithm pioneered by *Page* and *Brin* [6.59]. Using PageRank, the "importance" (here consisting of reputation, value, and impact) of a scholar depends not only on the number of scholars that vote for her, but also their importance. The more important the scholars that link to a person, the more important the person must be. The FundRank model was validated using citation data from 37 million papers over 20 years as a proxy for how each scientist might distribute funds within the proposed system. Simulation results show funding patterns that have a similar distribution compared to NSF and NIH funding for the past decade—at a fraction of the cost required by the current system.

6.5 Challenges

Using mathematical, statistical, and computational models of STI in decision making poses a new and diverse set of challenges, many of which can be viewed as opportunities. Such challenges/opportunities are related to fundamental research, applied research, cyberinfrastructure, education, and outreach.

6.5.1 Fundamental Research

Research concerning STI is conducted across a wide range of disciplines, including (but not limited to): economics, social sciences, information sciences, science policy, scientometrics/bibliometrics, and physics. Researchers in these disciplines develop mathematical, statistical, and computational models of different types (stochastic, agent-based, epidemics, game-theoretic, network, etc.) to address the questions they are interested in.

One of the factors impeding the advancement of fundamental research is lack of free access to high-quality data. Such access would significantly reduce data curation efforts currently being done by each individual team, and as a consequence, would enable reproducibility. An additional challenge facing this type of research is the lack of obvious sources offering continuous funding.

Furthermore, researchers exploring STI modeling tend to publish in a wide range of venues, often addressing vastly different audiences. Current research efforts and the results of said efforts are not universally known to the researchers, let alone policy makers. Such a widespread state of ignorance slows scientific progress and can even lead to unnecessary reinventions of the wheel. Scientific events that foster interactions among intellectually diverse communities and shed new light on problems by forcing researchers and practitioners to think and talk about their own research in new ways would help address this issue.

In the meantime, to arrive at policy-relevant solutions, researchers and analysts must pose good questions rather than focus solely on outcomes. Moving from descriptive to normative theories seems desirable. One major research challenge concerns the development of multiscale models—covering the micro (individual) to macro (population) levels—and understanding the appropriateness of particular models for particular scales. Ultimately, STI modeling experts should keep an open mind, and aim to learn from other branches of science (e. g., physics, economics, medicine) that are actively working on systems-science approaches.

6.5.2 Applied Research

One of the main reasons for the relatively low adoption rates of STI models is that these models are developed within different government institutions/agencies, and as a consequence often lack wider exposure. Relationships between model builders and users/stakeholders are often strained by poor communication at all stages of development, from the initial design (what question is being asked, what assumptions are being made, what measures and metrics are being used, etc.) to the interpretation and application of the results to real-world problems. This strain is further exasperated by an inherently opaque modeling process that neither creates nor maintains a sense of *buy-in* from the very beginning of a project.

For those interested in further reading, there are a few case studies that provide insights into the possibilities and challenges of carrying out applied research using modeling [6.60]: *Charles Phelps* et al., for example, implemented a tool for the measurement of the importance of vaccines—*SMART Vaccines*—which was then used by decision makers rather than model builders [6.61].

6.5.3 Cyberinfrastructure

As with many other disciplines, a robust cyberinfrastructure (e. g., data and model repositories, computing and visualization infrastructures) will greatly benefit STI modeling efforts. Many of the sciences have already setup billion-dollar international data infrastructures, and distributed computing systems in close collaboration with their government and industry partners, with impressive effect. Such synergy can be seen in the fields of meteorology (e. g., weather forecasts and hurricane and tornado prediction), epidemiology (e. g., predicting the next pandemic and identifying the best intervention strategies), climate research (e. g., predicting future scenarios and setting carbon prices), and financial engineering (e. g., stock trading and pricing predictions).

Sadly, no such universal infrastructure yet exists for the study and management of STI modeling, leading to up to 80% of project efforts being commonly spent on the acquisition, cleaning, interlinkage, and preprocessing of relevant data. Despite great benefits of building common infrastructure that we've witnessed in the natural sciences—where building a general infrastructure of commonly used data available to all has led to major advances (e. g., climate studies, astronomy, etc.)—STI modeling resources have been largely spent on individual project levels. Such a model of funding is unconducive to quick advancement of this area; successful STI modeling requires validation, iterative improvement, and a community of users, all of which could be provided via appropriate cyberinfrastructure. However, building such an infrastructure will require active partnerships among academia, government, and industry.

6.5.4 Education and Outreach

Advancing science, technology, and innovation requires extensive education and training. Recent studies show that data visualization literacy—the ability to read and write data visualizations—is relatively low [6.62]. Going forward, introducing computational modeling into formal and informal education will prove vital. More proactive and involved partnerships between stakeholders and modelers will allow for simpler models that can be understood and validated more easily. Such active partnerships will in turn help modelers deliver a timely and effective product, while also helping stakeholders determine their usefulness. At the same time, there is an urgent need for researchers, model builders and other users to enhance their communication and visualization skills.

Modeling results also need to be communicated effectively to different types of stakeholders. Storytelling and the art of communicating major results and recommendations in a clear and simple message is vital. Recent reports by the US National Academy of Sciences [6.63] and the National Academies of Sciences, Engineering, and Medicine [6.64] emphasize the importance of communication with nonscientists, and provide excellent examples of how such communication can be achieved.

6.6 Insights and Opportunities

As we have made clear, computational models of STI are deeply complex, and special effort is required to communicate not only their inner workings, but the implications of their results to relevant stakeholders. With this in mind, visualizations of data quality, data analysis, model parameter effects, and near real-time forecasts of STI developments can substantially increase the adoption rate and utility of modeling efforts.

6.6.1 Modeling Needs and Implementation

Modeling research and development strongly depend on understanding the problem at hand, as well as the range of actions a decision maker can take in response to that problem. If the wrong problem is modeled, or if suggested actions are infeasible (e. g., doubling the US R&D funding budget), then model utility as a consequence will be low.

Furthermore, there is a major difference between statistical significance and *business relevance*. For example, models used by PayPal need to avoid causing substantial costs via *false positives* (unidentified malicious users that cost PayPal money) but also via *false negatives* (valued customers with blocked accounts that cost PayPal reputation and might lead to bad press).

Model validation is of paramount importance. Ideally, different types of models can be applied to capture the structure and dynamics of that very same complex system and only if multiple models predict the same results should these results be used to make informed decisions.

Experts will need to work across disciplinary and institutional boundaries to exploit synergies, and to arrive at modeling results that are greater than the sum of their parts. There is a need for—and advantage to be gained from—combining basic and applied contract work [6.65]. Model developers (e. g., in academia and industry) should aim to *room in* with model users (policy and other decision makers) in an effort to foster active relationships.

Computational models also need to be vetted by experts, and as a consequence earn the trust of the scientific policymaking community before many start using them in practice. Key to building trust is the dogged pursuit of transparency, and also engaging stakeholders in the design and application of STI models. Easy-to-use, simple models that answer real-world questions are more readily adopted by decision makers than complex models with many parameter values.

Different policy offices have different abilities to absorb and implement models. Resistance to the adoption of new tools and approaches in general is unavoidable; the United States Federal Government, perhaps most notoriously, is the largest and most complex organization in the world, yet remains poorly understood and continues to use outdated decision support tools and processes. Models could be extremely useful when making resource allocation decisions, whether promoting agency missions, or managing international crises. Systems dynamic modeling is considered the best option, and yet not much has changed over the last decade since these approaches were first suggested. This can be attributed—simply, and frustratingly—to the human unwillingness to adapt and change.

6.6.2 Data Infrastructure

High-quality and high-coverage data is an imperative ingredient in high-quality modeling results. Currently, multiple teams are overlapping their efforts in cleaning, interlinking, and processing the same data (e. g., publication or patent data), and as a consequence are reducing the total amount of resources that could be spent on model research or validation. What's worse is that such data is preprocessed in slightly different ways across teams, making it hard or impossible to replicate results across sites.

While having so-called *big data* regarding science and technology dynamics is important to answer certain questions, having *more data* is not—and should not be—the answer to modeling questions. Though a large number of modelers use unstructured data, structured data boasts unique values, and as it becomes increasingly available [6.66] should also be explored. Given that many high-quality datasets are held by various sectors of industry (e. g., Web of Science and Scopus publication data, LinkedIn expertise profile data, Twitter or Instagram data, etc.) it appears highly desirable to work closely with them. Going forward, data sharing, data repositories, and joint data curation efforts should be explored as universal practices.

6.6.3 Code Repository and Standards

Efficient means by which to share STI model code are essential, not only to ensure replicability and reproducibility of model results, but also to support model comparisons and effective teaching. While some teams actively use the repository GitHub.com to share code and documentation, STI models remain difficult to locate among the millions of open-source code projects stored there.

The time is ripe to focus the energies and resources of researchers on building a cyberinfrastructure and a research community to support both systematic research and development efforts. Instead of creating a new repository, it would be most efficient to build upon and extend/interlink existing model repositories. Existing data repositories can be broken into three categories: academic, government, and industry.

Academic Repositories

Academic repositories are typically associated with a tool. For example:

- *Agent Modeling Platform* (AMP) project provides "extensible frameworks and exemplary tools for representing, editing, generating, executing and visualizing agent-based models (ABMs) and any other domain requiring spatial, behavioral and functional features." (http://www.eclipse.org/amp)
- *GAMA* is a "modeling and simulation development environment for building spatially explicit agent-based simulations." (https://github.com/gama-platform)
- *NetLogo* is a "multi-agent programmable modeling environment. It is used by tens of thousands of students, teachers and researchers worldwide. It also powers HubNet participatory simulations." (http://ccl.northwestern.edu/netlogo)
- *MASON* is a "fast discrete-event multi-agent simulation library core in Java, designed to be the foundation for large custom-purpose Java simulations, and also to provide more than enough functionality for many lightweight simulation needs. MASON contains both a model library and an optional suite of visualization tools in 2D and 3D." (http://cs.gmu.edu/~eclab/projects/mason)
- The *Repast Suite* is a "family of advanced, free, and open source agent-based modeling and simulation platforms that have collectively been under continuous development for over 15 years." (http://repast.sourceforge.net)

Repositories might also be created for specific research projects. For example, SIMIAN (http://www.simian.ac.uk) funded by the Economic and Social Research Council to promote and develop social simulation in the UK, uses the SKIN model (https://github.com/InnovationNetworks/skin). Another example is OpenABM (https://www.openabm.org) that provides a growing collection of tutorials and FAQs on agent-based modeling as part of the CoMSES Network.

Government Institutions

Government institutions aim to support sharing of datasets or tools. NSF's *SciSIP* program maintains a listing of "Datasets, Graphics & Tools" pertinent to the *Science of Science Policy* (SOSP) community at http://www.scienceofsciencepolicy.net/datasets_tools.

The *Interagency Modeling and Analysis Group* (IMAG) (https://www.imagwiki.nibib.nih.gov) and the Multiscale Modeling Consortium aim to grow the field of multiscale modeling in biomedical, biological and behavioral systems, to promote model sharing and the development of reusable multiscale models, and disseminate the models and insights gained from the models to the larger biomedical, biological, and behavioral research community, among others.

The *Predictive Model Index* lists over 100 reusable, sharable models in support of reproducible science (https://www.imagwiki.nibib.nih.gov/model-indexing). The *Centers for Disease Control and Prevention* (CDC) made the "H1N1 Flu (Swine Flu): Preparedness Tools for Professionals" software available at http://www.cdc.gov/h1n1flu/tools. The page was developed during the 2009–2010 H1N1 pandemic, but it has not been updated, and is being archived for historic and reference purposes only.

Publishers typically aim to ensure replicability of work by asking authors to submit datasets and models. Examples are The *Journal of Artificial Societies and Social Simulation* (JASSS, http://jasss.soc.surrey.ac.uk/JASSS.html), an interdisciplinary journal for the exploration and understanding of social processes by means of computer simulation; published since 1998, *JASSS* recommends authors upload model code and associated documentation to the CoMSES Net Computational Model Library (https://www.comses.net/codebases/). As of June 2016, the CoMSES library featured 352 agent-based models.

Industry

Industry has long embraced big data and advanced data mining, modeling, and visualization algorithms. Computational models are widely used in online recommendation services (e. g., those provided by Amazon or Netflix), and by financial and insurance companies (e. g., to detect credit card fraud and estimate fees). Many companies use models internally to support strategic decision making, and to guide investment decisions. While code is typically proprietary, close industry-academia-government collaborations are likely beneficial for all parties involved.

6.6.4 Visualization and Communication of Modeling Results

Global operation rooms that provide visualizations of current data and predictions of possible futures (already commonplace in the fields of meteorology, finance, epidemiology, and defense) might soon be commonplace in support of funding, strategic intelligence, or policy decision making.

William Rouse has been pioneering "policy flight simulators" that let decision makers fly the future before they write the check [6.10]. His team uses a combination of commercial off-the-shelf tools (e. g., AnyLogic, D-3, Excel, R, Simio, Tableau, and Vensim) rather than writing software from scratch. This practice can enable creation of a prototype interactive environment within a week or two, which in turn allows rapid user feedback and easy midcourse corrections.

Meanwhile, *Ben Shneiderman* and his team developed EventFlow, a novel tool for event sequence analytics that includes a timeline display showing all individual records, their point and interval events, as well as an aggregated view of all the sequences in the dataset (http://hcil.umd.edu/eventflow) [6.67]. Among others, the tool supports the examination of data quality before any type of data analysis is conducted or visualizations are rendered—blind usage of data is dangerous.

Storytelling in particular provides a powerful means to communicate data analysis and modeling results [6.63, 68]. Merging data with narrative, especially when communicating the value of research, is a primary way to connect to policymakers.

Katy Börner and her team are developing and prototyping *Science Forecasts*; a news show that communicates local and global developments in science, technology, and innovation to a general audience. In Spring 2015, a pilot episode was recorded featuring a moderator that explained trends using an animated map of science, analogous to a weather forecast. Zeroing in on specific research results using Twitter for detecting episodes of depression, the information was presented by Johan Bollen and Fred Cate, both faculty at Indiana University. A still image of the news can be seen in Fig. 6.2.

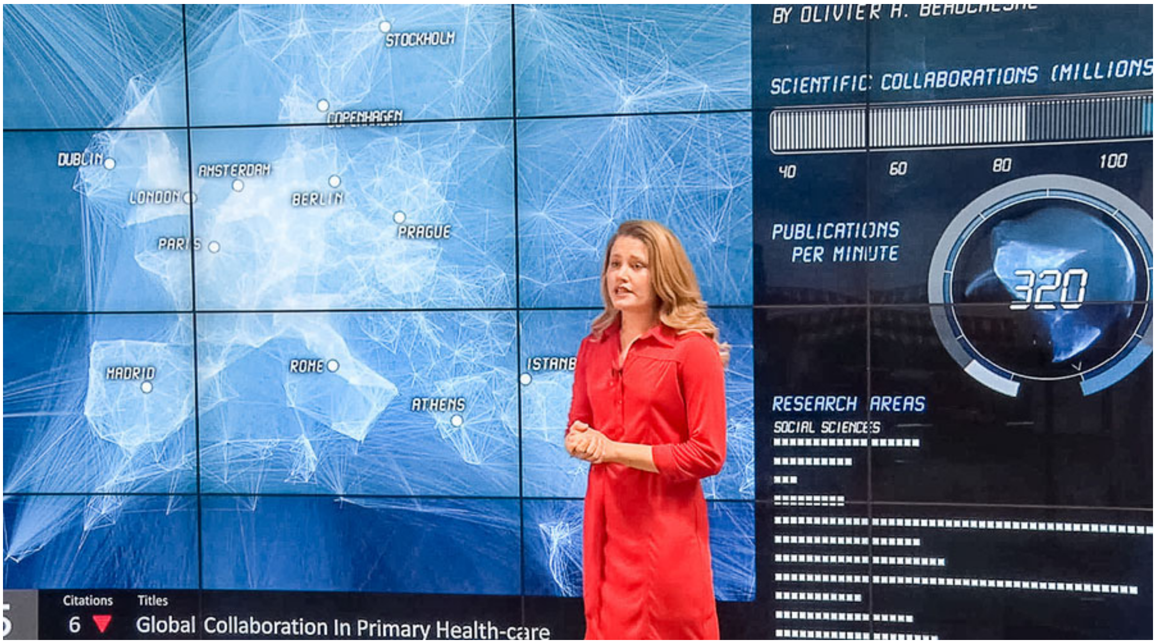


Fig. 6.2 *Science Forecast*, recorded at Indiana University, presents interviews and animated maps of scientific activity in a manner similar to weather forecasts. The program demonstrates the power of data and visual analytics to provide up-to-date stories on science trends and developments

6.7 Outlook

In 2007, *Issues in Science and Technology* published *The Promise of Data-Driven Policymaking* by *Daniel Esty* and *Reece Rushing* [6.69]. In 2016, the same magazine published "Data-Driven Science Policy" [6.7]. The articles both point out that in the corporate sector, a wide variety of data-driven approaches are used to boost profits, including systems that improve performance and reliability, evaluate the success of advertising campaigns, and determine optimal pricing. Both articles argue for the need for—and discuss the premise of—data-driven decision making and policy making in STI using large-scale, high-quality datasets, and computational means to inform human decision makers.

Today, in 2019, a wide range of mathematical, statistical, and computational models exist that were developed and implemented in a variety of settings to increase our collective understanding of the structure and dynamics of STI and to support human decision making. While academic researchers typically focus on work that can be published, there is a growing emphasis by researchers and practitioners on the power of models to advance future decision making, and to communicate the usefulness of models to simulate, explain, and communicate the past, present, and future.

Acknowledgments. The work was supported in part by the National Institutes of Health under awards U01CA198934, P01AG0393, and OT2OD026671 and National Science Foundation awards 1546824, 1713567, 1735095, and 1839167, NETE Federal IT, Thomson Reuters, Indiana University Network Science Institute, and the Cyberinfrastructure for Network Science Center at Indiana University. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

6.1 P. Ahrweiler, N. Gilbert, A. Pyka (Eds.): *Joining Complexity Science and Social Simulation for Innovation Policy: Agent-based Modelling using the SKIN Platform* (Cambridge Scholars, Newcastle upon Tyne 2015)
6.2 C. Watts, N. Gilbert: *Simulating Innovation. Computer-Based Tools for Re-Thinking Innovation* (Edward Elgar, London 2014)
6.3 D. Hicks, P. Wouters, L. Waltman, S. de Rijcke, I. Rafols: The Leiden Manifesto for research metrics, Nature **520**, 430–431 (2015)
6.4 D.J. de Solla Price: *Little Science, Big Science* (Columbia Univ. Press, New York 1963)
6.5 S. Milojević: Principles of scientific research team formation and evolution, Proc. Natl. Acad. Sci. USA **111**(11), 3984–3989 (2014)
6.6 S. Wuchty, B.F. Jones, B. Uzzi: The increasing dominance of teams in production of knowledge, Science **316**(5827), 1036–1039 (2007)
6.7 K. Börner: Data-driven science policy, Issues Sci. Technol. **32**(3), 26–28 (2016)
6.8 A. Scharnhorst, K. Börner, P. van den Besselaar (Eds.): *Models of Science Dynamics: Encounters between Complexity Theory and Information Science* (Springer, Berlin 2012)
6.9 J. de Rosnay: *The Macroscope: A New World Scientific System* (Harper Row, New York 1979)
6.10 W.B. Rouse: Human interaction with policy flight simulators, J. Appl. Ergon. **45**(1), 72–77 (2014)
6.11 W.B. Rouse: *Modeling and Visualization of Complex Systems and Enterprises: Explorations of Physical, Human, Economic, and Social Phenomena* (Wiley, Hoboken 2015)
6.12 P. Ahrweiler, M. Schilperoord, A. Pyka, N. Gilbert: Modelling research policy: Ex-ante evaluation of complex policy instruments, J. Artif. Soc. Soc. Simul. **18**(4), 5 (2015)
6.13 C.A. Lave, J.G. March: *An Introduction to Models in the Social Sciences* (Univ. Press of America, Lanham 1993)
6.14 J.H. Miller, S.E. Page: *Complex Adaptive Systems: An Introduction to Computational Models of Social Life* (Princeton Univ. Press, Princeton 2007)
6.15 P. Cilliers: *Complexity and Postmodernism: Understanding Complex Systems* (Routledge, London 1998)
6.16 G. Nicolis, I. Prigogine: *Exploring Complexity: An Introduction* (W.H. Freeman Company, New York 1989)
6.17 J.H. Holland, K.J. Holyoak, R.E. Nisbett, P.R. Thagard: *Induction: Processes of Inference, Learning, and Discovery* (MIT Press, Cambridge 1986)
6.18 E. Winsberg: *Science in the Age of Computer Simulation* (Univ. Chicago Press, Chicago 2010)
6.19 M. Morrison: *Reconstructing Reality: Models, Mathematics, and Simulations* (Oxford Univ. Press, Oxford 2015)
6.20 P.N. Edwards: *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming* (MIT Press, Cambridge 2010)
6.21 J.M. Epstein: *Generative Social Sciences: Studies in Agent-Based Computational Modeling* (Princeton Univ. Press, Princeton 2006)
6.22 N. Gilbert: *Agent-based Models* (SAGE, Los Angeles 2008)
6.23 N. Gilbert, K.G. Troitzsch: *Simulation for the Social Scientist*, 2nd edn. (Open Univ. Press, Maidenhead 2009)

6.24 T. Hastie, R. Tibshirani, J. Friedman: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edn. (Springer, New York 2017)
6.25 K.P. Murphy: *Machine Learning: A Probabilistic Perspective* (MIT Press, Cambridge 2012)
6.26 J. Kleinberg, J. Ludwig, S. Mullainathan, Z. Obermeyer: Prediction policy problems, Am. Econ. Rev. Papers Proc. **105**(5), 491–495 (2015)
6.27 K. Börner, W. Glänzel, A. Scharnhorst, P. van den Besselaar: Modeling science: Studying the structure and dynamics of science, Scientometrics **89**(1), 346–463 (2011)
6.28 K. Börner, B. Edmonds, S. Milojević, A. Scharnhorst: Simulating the processes of science, technology, and innovation, Scientometrics **110**(1), 385 (2016)
6.29 K. Börner, K.W. Boyack, S. Milojević, S.A. Morris: An introduction to modeling science: Basic model types, key definitions, and a general framework for comparison of process models. In: *Models of Science Dynamics: Encounters Between Complexity Theory and Information Sciences*, ed. by A. Scharnhorst, K. Börner, P. van den Besselaar (Springer, Berlin 2012) pp. 3–22
6.30 J. Smith, C. Jenks: *Qualitative Complexity: Ecology, Cognitive Processes and the Re-Emergence of Structures in Post-Humanist Social Theory* (Routledge, London 2006)
6.31 E.C.M. Noyons, A.F.J. van Raan: Monitoring scientific developments from a dynamic perspective: Self-organized structuring to map neural network research, J. Am. Soc. Inf. Sci. **49**(1), 68–81 (1998)
6.32 A.F.J. van Raan: Fractal dimension of co-citations, Nature **347**, 626 (1990)
6.33 A.F.J. van Raan: On growth, ageing, and fractal differentiation of science, Scientometrics **47**(2), 347–362 (2000)
6.34 W.O. Kermack, A.G. McKendrick: A contribution to the mathematical theory of epidemics, Proc. R. Soc. A **115**, 700–721 (1927)
6.35 W. Goffman: Mathematical approach to the spread of scientific ideas – The history of mast cell research, Nature **212**(5061), 449–452 (1966)
6.36 D.J. de Solla Price: Networks of scientific papers, Science **149**, 510–515 (1965)
6.37 D.J. de Solla Price: A general theory of bibliometric and other cumulative advantage processes, J. Am. Soc. Inf. Sci. **27**(5), 292–306 (1976)
6.38 F. Radicchi, S. Fortunato, C. Castellano: Universality of citation distributions: Toward an objective measure of scientific impact, Proc. Natl. Acad. Sci. USA **105**, 17268–17272 (2008)
6.39 Y.-H. Eom, S. Fortunato: Characterizing and modeling citation dynamics, PLoS ONE **6**(9), e24926 (2011)
6.40 P.D.B. Parolo, R.K. Pan, R. Ghosh, B.A. Huberman, K. Kaski, S. Fortunato: Attention decay in science, J. Informetrics **9**(4), 734–745 (2015)
6.41 D. Wang, C. Song, A.-L. Barabási: Quantifying long-term scientific impact, Science **342**(6154), 127–132 (2013)
6.42 R. Klavans, K.W. Boyack: Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge?, J. Assoc. Inf. Sci. Technol. **68**, 984–998 (2016)
6.43 T. Kuhn, M. Perc, D. Helbing: Inheritance patterns in citation networks reveal scientific memes, Phys. Rev. X **4**, 041036 (2014)
6.44 F. Shi, J.G. Foster, J.A. Evans: Weaving the fabric of science: Dynamic network models of science's unfolding structure, Soc. Netw. **43**, 73–85 (2015)
6.45 A.M. Petersen, M. Riccaboni, H.E. Stenley, F. Pammolli: Persistence and uncertainty in the academic career, Proc. Natl. Acad. Sci. USA **109**, 5213–5218 (2012)
6.46 R. Sinatra, D. Wang, P. Deville, C. Song, A.-L. Barabási: Quantifying the evolution of individual scientific impact, Science **354**(6312), aaf5239 (2016)
6.47 N. Payette: Agent-based models of science. In: *Models of Science Dynamics: Encounters Between Complexity Theory and Information Sciences*, ed. by A. Scharnhorst, K. Börner, P. van den Besselaar (Springer, Berlin 2012) pp. 127–157
6.48 N. Gilbert: A simulation of the structure of academic science, Sociol. Res. Online (1997), https://doi.org/10.5153/sro.85
6.49 K. Börner, J. Maru, R. Goldstone: The simultaneous evolution of author and paper networks, Proc. Natl. Acad. Sci. USA **101**, 5266–5273 (2004)
6.50 A.-L. Barabási, R. Albert: Emergence of scaling in random networks, Science **286**, 509–512 (1999)
6.51 R.K. Merton: The Matthew effect in science, Science **159**(3810), 56–63 (1968)
6.52 X. Sun, J. Kaur, S. Milojević, A. Flammini, F. Menczer: Social dynamics of science, Sci. Rep. **3**, 1069 (2013)
6.53 A. Clauset, D.B. Larremore, R. Sinatra: Data-driven predictions in the science of science, Science **355**, 477–480 (2017)
6.54 P. Azoulay: Turn the scientific method on ourselves, Nature **483**, 31–32 (2012)
6.55 P. Azoulay, J.G. Zivin, G. Manso: Incentives and creativity: Evidence from academic life sciences, RAND J. Econ. **42**(3), 527–554 (2011)
6.56 O.A.D. Arrieta, F. Pammolli, A.M. Petersen: Quantifying the negative impact of brain drain on the integration of European science, Sci. Adv. **3**, e1602232 (2017)
6.57 J. Kleinberg, H. Lakkaraju, J. Leskovec, J. Ludwig, S. Mullainathan: Human decisions and machine predictions, NBER Working Paper No. 23180 (2017)
6.58 J. Bollen, D. Crandall, D. Junk, Y. Ding, K. Börner: From funding agencies to scientific agency: Collective allocation of science funding as an alternative to peer review, EMBO Reports **15**(2), 131–133 (2014)
6.59 L. Page, S. Brin: The anatomy of a large-scale hypertextual web search engine, Comput. Netw. and ISDN Syst. **30**(1–7), 107–117 (1998)
6.60 K. Börner, S. Milojević (Eds.): *Modeling Science, Technology and Innovation. NSF Conference Report, Indiana University*, https://modsti.cns.iu.edu/report and presenter slides at https://modsti.cns.iu.edu (2016)
6.61 C. Phelps, G. Madhavan, K. Sangha, R. Rappuoli, R.R. Colwell, R.M. Martinez, L. King: A priority-setting aid for new vaccine candidates, Proc. Natl. Acad. Sci. USA **111**(9), 3199–3200 (2014)

6.62 K. Börner, J.E. Heimlich, R. Balliet, A.V. Maltese: Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors, Inf. Vis. **15**(3), 198–213 (2016)
6.63 US National Academy of Sciences: *The Science of Science Communication II: Summary of a Colloquium* Retrieved from Washington (2014)
6.64 National Academies of Sciences, Engineering, and Medicine: *Communicating Science Effectively: A Research Agenda* Retrieved from Washington (2017)
6.65 B. Shneiderman: *The New ABCs of Research: Achieving Breakthrough Collaborations* (Oxford Univ. Press, Oxford 2016)
6.66 J. Hendler, W. Hall: Science of the World Wide Web, Science **354**(6313), 703–704 (2016)
6.67 M. Monroe, R. Lan, C. Plaisant, B. Shneiderman: Temporal event sequence simplification, IEEE Trans. Vis. Comput. Graph. **19**(12), 2227–2236 (2013)
6.68 E. Segel, J. Heer: Narrative visualization: Telling stories with data, IEEE Trans. Vis. Comput. Graph. **16**(6), 1139–1148 (2010)
6.69 D. Esty, R. Rushing: The promise of data-driven policymaking, Issues Sci. Technol. **23**(4), 67–72 (2007)

Katy Börner

School of Informatics, Computing, and Engineering
Indiana University
Bloomington, IN, USA
katy@indiana.edu

Katy Börner is Victor H. Yngve Distinguished Professor of Engineering and Information Science in the School of Informatics, Computing, and Engineering, Adjunct Professor at the Department of Statistics in the College of Arts and Sciences, Core Faculty of Cognitive Science at Indiana University. She holds an MS in Electrical Engineering from the University of Technology in Leipzig (1991) and a PhD in Computer Science from the University of Kaiserslautern (1997).

Staša Milojević

School of Informatics, Computing, and Engineering
Indiana University
Bloomington, IN, USA
smilojev@indiana.edu

Staša Milojević is an Associate Professor of Informatics in the Center for Complex Network and Systems Research in the School of Informatics, Computing, and Engineering at Indiana University. Her work covers a range of topics within the "science of science": dynamics of research teams, collaborative networks, formation and evolution of scientific fields, and research metrics. She received a PhD in Information Studies from the University of California, Los Angeles.

7. Science Mapping Analysis Software Tools: A Review

Jose A. Moral-Munoz, Antonio G. López-Herrera, Enrique Herrera-Viedma, Manuel J. Cobo

Scientific articles are one of the most important types of output of a researcher. In that sense, bibliometrics is an essential tool for assessing and analyzing academic research output contributing to the progress of science in many different ways. It provides objective criteria to assess research developed by researchers, being increasingly valued as a tool for measuring scholarly quality and productivity. Science mapping is a bibliometric tool to analyze and mine scientific output. The aim of this chapter is to present a thorough review of science mapping software tools, showing strengths and limitations. Six software tools that meet the criteria of being free, full, and allowing the whole analysis to be performed are analyzed:

- BibExcel
- CiteSpace II
- CitNetExplorer
- SciMAT
- Sci^2 Tool
- VOSviewer.

This analysis describes aspects related to data processing, analysis options, and visualization. The particular properties of each tool that allows us to analyze the science are presented, the choice of a particular tool one depends on the type of actor to be analyzed and the output expected.

7.1 Science Mapping Analysis

According to *Price* [7.1], science, in a concise definition, is determined as that which is published in scientific publications, and the researcher is the person who collaborated in writing some of these publications. Therefore, it is clear that the publication is the mean by which research is evaluated, validated, and transferred. With respect to the set of scientific publications, attention is mainly focused on documents published on academic journals, but there are also other kinds of publications, such as proceedings, patents, books, etc. As mentioned in the above definition, academic articles published by researchers have developed at quick pace. Moreover, nowadays there are many global bibliographic databases, indexing millions of scientific documents. For example, the Web of Science (WoS) core collection indexes more than 12 000 journals, and Scopus indexes around 18 000 journals. These kinds of databases offer a wealth of information (such as authors, titles, journals, keywords, institutions, citations, etc.) to be analyzed, and consequently, generate new knowledge in relation to different aspects of the science.

In order to analyze and evaluate scientific activity, a huge amount of data is needed. Thus, the analysis of this data is performed using intelligent techniques and tools. In this sense, bibliometrics is an important tool for assessing and analyzing the scientific production contributing to the advance of science in a wide range of ways [7.2, 3]. Furthermore, it gives objective criteria to evaluate a researcher's production, being progressively valued as a tool for assessing academic

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_7

quality and productivity [7.4]. It is a useful approach to assess and analyze the different actors that produce research [7.5]:

1. Countries
2. Universities
3. Research centers
4. Research groups
5. Journals
6. Researchers.

On the other hand, it should be noted that bibliometrics has been used for quantitative production evaluation, but in the last decades it has been starting to be used as a research tool in its own right [7.6, 7]. The approaches employed to perform the bibliometric analyses can be used to explore the impact of a research area, a set of researchers, or a particular article [7.8]. In addition, the analysis output could help to optimize research allocation, reorient funding, rationalize organizations, and restrict or increase productivity in other areas [7.9].

In relation to bibliometrics, it can be divided into two main areas [7.10]: performance analysis and science mapping analysis. The former assesses the different scientific producers using the bibliographic data and applying bibliometric indexes (e. g., *H*-index [7.11], HG-index [7.12], etc.). The latter is dedicated to showing the structural and dynamic aspects of a research field, and how it evolves through time [7.13–15]. Particularly, it tries to find a representation of the intellectual connections and its evolution in a knowledge area [7.16].

In order to perform a science mapping analysis, a great variety of software tools can be used. Among the available tools, some have a generic aim and come from the field of statistics or social network analysis. Nevertheless, in 2011, *Cobo* et al. [7.17] described the main software tools developed specifically for this task. Since then, new software tools have appeared, and new versions of the previously described tools have been released. Thus, an updated review of the available software is given in this chapter. The tools that meet the following criteria have been taken into account:

1. They are freely available.
2. The software is complete in itself.
3. They allow to perform the whole analysis to be performed.

Nevertheless, there are other interesting options to cover different necessities or preferences.

Regarding generic software, Pajek [7.18] and Gephi [7.19] are the most remarkable options. They allow analysis of different aspects related to networks and offer several visualization modes. Nonetheless, they do not have preprocessing tools that allow to perform some transformations from the retrieved data. In other words, the user may use the final network file.

Among the APIs (application programming interface), there are also different alternatives available, such as Bibliometrix [7.20], Citan [7.21], Sciento-text [7.22], and BiblioTools [7.23]. The three first tools are based on R packages and BiblioTools in Python. They offer different analysis options to perform different analyses but have the inconvenience using non-bibliometric full software.

Other available tools are those based on a relational database model in which the user can perform some Structured Query Language (SQL) queries [7.24, 25]. These kinds of tools are designed to perform science mapping analysis, extracting different indicators.

Furthermore, there are others tools focused on preprocessing step, such as Software Tool for Improving and Converting Citation Indices (STICCI) [7.26]. It performs several dataset format conversions, corrects mistakes, duplications, misspellings, name variants, etc. Thus, it is an interesting tool for preparing the database.

On the other hand, it is worth highlighting a commercial option, VantagePoint [7.27]. This is a tool for discovering knowledge in relation to patent and academic databases.

In addition to software focused on science mapping analysis, there are other tools that allow performing other kinds of bibliometric studies, such as CRExplorer [7.28]. This is focused on detecting the main production of a field, based on the referenced publication years spectroscopy (RPYS) method [7.29].

The chapter is structured in four main sections. In the first section, the main characteristics of the bibliometric networks are described. In the second section, each software tool is thoroughly described, highlighting the distinctive features of each one. In the third section, a comparative analysis among the software is shown. In the last section, some conclusions from the review performed are made. Thereupon, the aim of this chapter is to introduce readers to the science mapping analysis tools available.

7.2 Bibliographic Networks

An important feature in relation to the available science mapping analysis software is the possibility of building several networks according to the different units of analysis. In other words, the software capacity to extract different bibliometric networks is of central importance.

Different bibliographic or bibliometric networks are available in the scientific literature [7.30, 31]. The main characteristic of these types of networks is that they are composed of nodes and edges. Nodes represent publications, journals, researchers, or words. Edges show relations between two nodes. Bibliometric networks used are weighted, and the most common are based on relations of citations, keyword co-occurrence, and co-authorship. In the networks, the edges not only indicate the relation between the nodes, also the strength of the relation. In the following paragraphs, the co-citation, bibliographic coupling, keywords co-occurrences, and co-authorship networks will be briefly described. Table 7.1 shows the characteristics of the main bibliometric networks [7.17]:

1. Bibliometric techniques
2. Units of analysis
3. Type of relations analyzed by these techniques:

- Co-citation networks: According to the definition of co-citation [7.32], two publications can be considered as being co-cited if a third publication cites both of them. The strength of the co-citation relation will depend on the number of publications that cites both publications together. Thus, the higher the number of co-citations, the stronger the co-citation relation.
- Bibliographic coupling networks: In this case, two publications are bibliographically coupled if a third publication is cited by both publications [7.33]. A higher number of references shared by two publications indicates a stronger bibliographic coupling relation between them.
- Co-authorship networks: In this kind of bibliometric network, the different actors (researchers, institutions, or countries) are linked to each other according to the number of publications that they have authored together [7.34].
- Word co-occurrences networks: In order to build this kind of network, terms are extracted from the title and abstract of a set of publications, as well as from the keywords listed by the author when the document was submitted. The number of co-occurrences of two keywords depends on the number of documents where both appear together in the title, abstract, or author's keyword list [7.35].

Table 7.1 Bibliometric networks types

Bibliometric techniques		Units of analysis	Relations
Co-citation	Author	Author's reference	Co-cited author
	Document	Reference	Co-cited documents
	Journal	Journal's reference	Co-cited journal
Bibliographic coupling	Author	Author's oeuvres	References among author's oeuvres
	Document	Document	References among documents
	Journal	Journal's oeuvres	References among journal's oeuvres
Co-author	Author	Author's name	Authors' co-occurrence
	Country	Country from affiliation	Countries' co-occurrence
	Institution	Institution from affiliation	Institutions' co-occurrence
Co-word		Keyword or term extracted from title, abstract or document's body	Terms' co-occurrence

7.3 Science Mapping Software

Several software tools have been specifically developed to analyze scientific domains by means of science mapping. In this chapter, the tools that meet the following criteria have been analyzed:

1. Freely available tools
2. Full software tools (APIs were excluded)
3. Tools that allow the whole analysis to be performed.

Therefore, we present six representative software that allow us to analyze different aspects in relation to science:

1. BibExcel [7.36]
2. CiteSpace II [7.37]
3. CitNetExplorer [7.38]
4. SciMAT [7.39]
5. Sci^2 Tool [7.40]
6. VOSviewer [7.41].

In Table 7.2, general information about the software analyzed is shown. Nevertheless, it should be taken into account that the software tools analyzed are still being developed, and, therefore, there will be new releases incorporating new features.

In the following sections, the main characteristics of these tools are depicted. Aspects related to data processing, analysis options, and visualization will be described for each tool.

7.3.1 BibExcel

BibExcel [7.36] is intended to help users analyze bibliographic information or any information of a textual nature formatted in a comparable manner. The idea is to create data files that can be imported to Excel or any program that takes tabbed data records (such as Statistical Package for the Social Sciences (SPSS), UCINET [7.42] or Pajek [7.18]) for further handling. It was developed by Olle Persson at the University of Umeå, Sweden.

This software incorporates various tools, some of them visible in the window and others hidden behind the menus. Many of the tools can be used as a part of mix to accomplish the desired outcome. The main characteristic of this software is its flexibility, but for this reason, it could be initially perceived as difficult to use by new users.

Furthermore, if we take a look to the recent research literature, several applied studies can be found. For instance, some of the latest studies carried out with BibExcel are focused in different thematic areas, such as:

1. Transport geography research [7.43]
2. Space research [7.44]
3. Intelligent transportation systems [7.45]
4. Scientific production on open access [7.46]
5. Ebola research [7.47]
6. Project management success [7.48].

Therefore, although this software does not offer the possibility to obtain a visualization without using a different tool, as has been observed, it is a tool frequently used in bibliometric studies.

Table 7.2 Software tools general information

Software tool	Analyzed version	Year	Developed by
BibExcel	2016-02-20	2016	University of Umeå (Sweden)
CiteSpace II	5.0.R4 SE	2017	Drexel University (USA)
CitNetExplorer	1.0.0	2014	Leiden University (The Netherlands)
SciMAT	1.1.04	2016	University of Granada (Spain)
Sci^2 Tool	1.2	2015	Cyberinfrastructure for Network Science Center (USA)
VOSviewer	1.6.5	2016	Leiden University (The Netherlands)

Data Processing

BibExcel can read information retrieved from various bibliographic sources, for example, WoS, Scopus, and the ProCite export format. Nevertheless, if the user has learned the typical BibExcel file structure, different types of documents can be formatted in order to be analyzed.

On the other hand, it allows us to submit the textual data to different preprocessing tasks, such as an English word stemmer, document deduplication, and text transformation (keeping authors' first initials, converting comma-delimited addresses, etc.). Besides, BibExcel empowers the deletion of low recurrence items and keeps just the strongest links. On the other hand, it provides different options to extract data (publication year, author name, etc.) and to remove it (DOI, reprint author, first column, etc.).

Analysis Options

Different bibliometric networks can be obtained using the parameters offered by this tool. The main networks are:

1. Co-citation
2. Bibliographic coupling
3. Co-author
4. Co-word.

Moreover, it is possible to create different co-occurrence matrices taking any document's field or combining some of them. Then, these matrices can be submitted to a normalization process using three measures:

1. Salton's cosine [7.49]
2. Jaccard's index [7.50]
3. Vladutz and Cook measures [7.51].

Once the data has been normalized, the user can apply a clustering algorithm or prepare a matrix to perform multidimensional scaling (MDS) using an external tool.

Visualization

In relation to the visualization, it does not present a specific tool for the output representation. It incorporates different export options that allow us to visualize the data using external software like Pajek [7.53], UCINET [7.42], SPSS, or VOSviewer [7.38]. The type of visualization selection will depend on the nature of the unit of analysis. Nevertheless, it offers the possibility of obtaining data files for interesting visualizations, such as the Google Maps (Fig. 7.1).

7.3.2 CiteSpace II

CiteSpace [7.37] is a Java application that allows the analysis and visualization of trends and patterns in a research area. The main goal of this tool is to facilitate the analysis of emerging trends in a knowledge domain. It was developed at Drexel University, USA.

This tool offers several options to understand and interpret network and historical patterns, such as the growth of a topic area, the main citations in the knowledge base, the automatic labeling of the different clusters using terms from citing articles, geospatial collaboration network, and international collaboration.

It is interesting to note that a previous version of this tool existed. It is now designed as CiteSpace II due to the major improvements performed. The main features that characterize this version are:

1. Burst detection
2. Betweenness centrality
3. Heterogeneous networks.

In order to show the great variety of studies performed with this tool, some of the latest research publications have been retrieved. The thematic areas studied are varied, such as:

1. Global rice [7.54]
2. Sustainable development [7.55]
3. Nonpoint source pollution [7.56]

Fig. 7.1 Google maps visualization using BibExcel data (from [7.52])

4. Molecular modeling in cyclodextrins [7.57]
5. Hospitality research [7.58].

In view of these publications, it is reasonable to think that this software is important and highly used in the applied scientific literature.

Data Processing

CiteSpace is able to work with different bibliographic formats, such as WoS, PubMed, arXiv, and the SAO/-NASA Astrophysics Data System (ADS). Moreover, analysis about grants and patents can be performed through National Science Foundation (NSF) awards and Derwent Innovations Index data.

In relation to data retrieval, CiteSpace allows us to obtain bibliographic records from PubMed directly from the tool. This function can be set to obtain a maximum of records. It is necessary to consider that these records do not include information on cited references, so it is not possible to perform a citation analysis. However, this data can be used to perform analyses about networks of collaboration, terms, keywords, and categories.

Analysis Options

According to the scientific literature [7.37], the CiteSpace workflow is briefly described in the following phases:

1. *Knowledge domain detection*. It is important to use the broadest possible term in order to cover the main available research in relation to an area.
2. *Data collection*. As stated above, this tool is able to use several bibliographic records.
3. *Extract research front terms*. It collects the terms from titles, abstracts, descriptors and identifiers of citing articles in datasets.
4. *Time slicing*. In this step, the user has to specify a specific time period in the whole interval and the length of the time slice selected.
5. *Threshold selection*. This threshold is based on three sets of levels for citation counts, co-citation counts, and co-citation coefficients.
6. *Pruning and merging*. By default, this tool use pathfinder network scaling [7.59]. Users can choose whether to apply this operation to individual networks. CiteSpace II implements a version of this algorithm that reduces the overall waiting time [7.37].
7. *Layout*. It supports a standard graph view an a timezone view.
8. *Visual inspection*. The tool allows the users to interact with the visualization in several ways. Some parameters can be controlled, such as attributes and labels.
9. *Pivotal points verification*. This step may be performed by an expert. The importance of the marked pivotal points has to be confirmed in order to obtain a correct final output.

Using CiteSpace, different types of bibliometrics networks can be obtained. The following networks can be constructed from the entered data:

1. Co-author [7.60, 61]
2. Co-author institutions
3. Co-author countries
4. Co-grants
5. Subject categories co-occurrence
6. Co-word [7.6, 35, 62]
7. Documents co-citation
8. Author co-citation
9. Journal co-citation [7.32, 63]
10. Documents bibliographic coupling [7.33, 64, 65].

On the other hand, the networks, or graphs obtained, can be analyzed by applying different time periods in order to detect the evolution of the domain studied. Moreover, the most important items from the network can be filtered by the analyst (e. g., the 50 most cited items for each time span). The matrices are normalized using three different measures:

1. Salton's cosine [7.49]
2. Dice [7.66]
3. The Jaccard index [7.50].

Furthermore, a burst-detection algorithm [7.67] is adapted to identify emergent research-font concepts. In order to highlight the potential pivotal points, Freeman's betweenness centrality metric [7.68] is used.

Visualization

CiteSpace II offers several possibilities to visualize and perform analysis from the networks built. The user obtains a visualization where the pivotal points are shown in relation to their betweenness centrality. They are highlighted in the software window with a ring surrounded by a tree ring (Fig. 7.2). This ring represents the citation history of an article. The color determines the time of corresponding citations, the thickness is proportional to the number of citations in a time slice, and finally, the number next to the node center is the number of citations received during the whole time period. In Fig. 7.3, an applied visualization in the scientometric area is shown.

Before obtaining the visualization, some characteristics can be selected, such as the static or animated mode

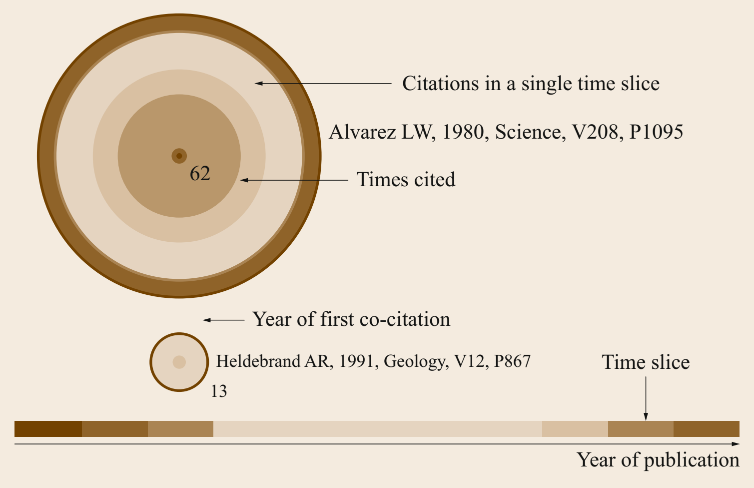


Fig. 7.2 CiteSpace II tree ring visualization (from [7.37])

and the type of networks. By default, it only shows the merged network, but it is possible to obtain the networks of all the time slices. These networks are opened in several extra windows. Once the visualization is shown, the user can interact, changing some features. For example, a persistent label can be set to a node, or two different nodes can be merged using the alias function.

Finally, it is interesting to highlight the geographical visualization option. Authors' geographic locations can be mapped as a geospatial map in keyhole markup language (KML). To obtain this representation, the Google Earth interface can be used (Fig. 7.4). Then, it is possible to explore authors' locations and links to their collaborators, and also to redirect to the original articles directly within Google Earth.

7.3.3 CitNetExplorer

CitNetExplorer is a software tool for visualizing and analyzing citation networks, based on the algorithmic historiography designed by *Garfield* [7.70]. In the networks obtained by CitNetExplorer, each node represents a publication. Each edge represents a citation relation between two publications. It was developed by the Centre for Science and Technology Studies (CWTS) at Leiden University, The Netherlands.

This software can be used for different purposes, such as analyzing the development of a research field over time, identifying the literature on a research topic, exploring the publication oeuvre of a researcher, or supporting literature reviewing [7.38]. Furthermore, not only does it focus on scientific publications, it may also be used to analyze other types of citation networks, such as patents.

In order to know the application of this software to carry out an analysis in different research areas, a search was conducted on the main bibliographic databases. CitNetExplorer has been used to analyze different areas in some recent publications:

1. Theme evolution of electrochemical energy storage research [7.71]
2. The journal *Nature of Nanotechnology* [7.72]
3. Successful aging [7.73]
4. Tribology [7.74]
5. Emerging induced pluripotent stem cell-based therapies for age-related macular degeneration [7.75].

Data Processing

The information to construct citation networks is collected directly from the WoS database. However, it is not restricted to this database, *Scopus* data could equally be analyzed. Then, CitNetExplorer is able to deal with large citation networks, including millions of publications and tens of millions of citation relations. Once the citation network has been formed, it can be exported into Pajek file format [7.53].

In order to construct the citation networks, CitNetExplorer uses two different approaches. One approach consists in providing both data from publications and data from citation relations among publications. The other approach consists in adding the data directly downloaded from WoS. In this case, in order to obtain the citation relations, CitNetExplorer is able to perform citation matching. There are two different citation matching functionality ways:

1. Based on DOI matching, but this data often is not available.
2. Based on first author's name, publication year, volume number, and page number. A perfect match is needed for each element.

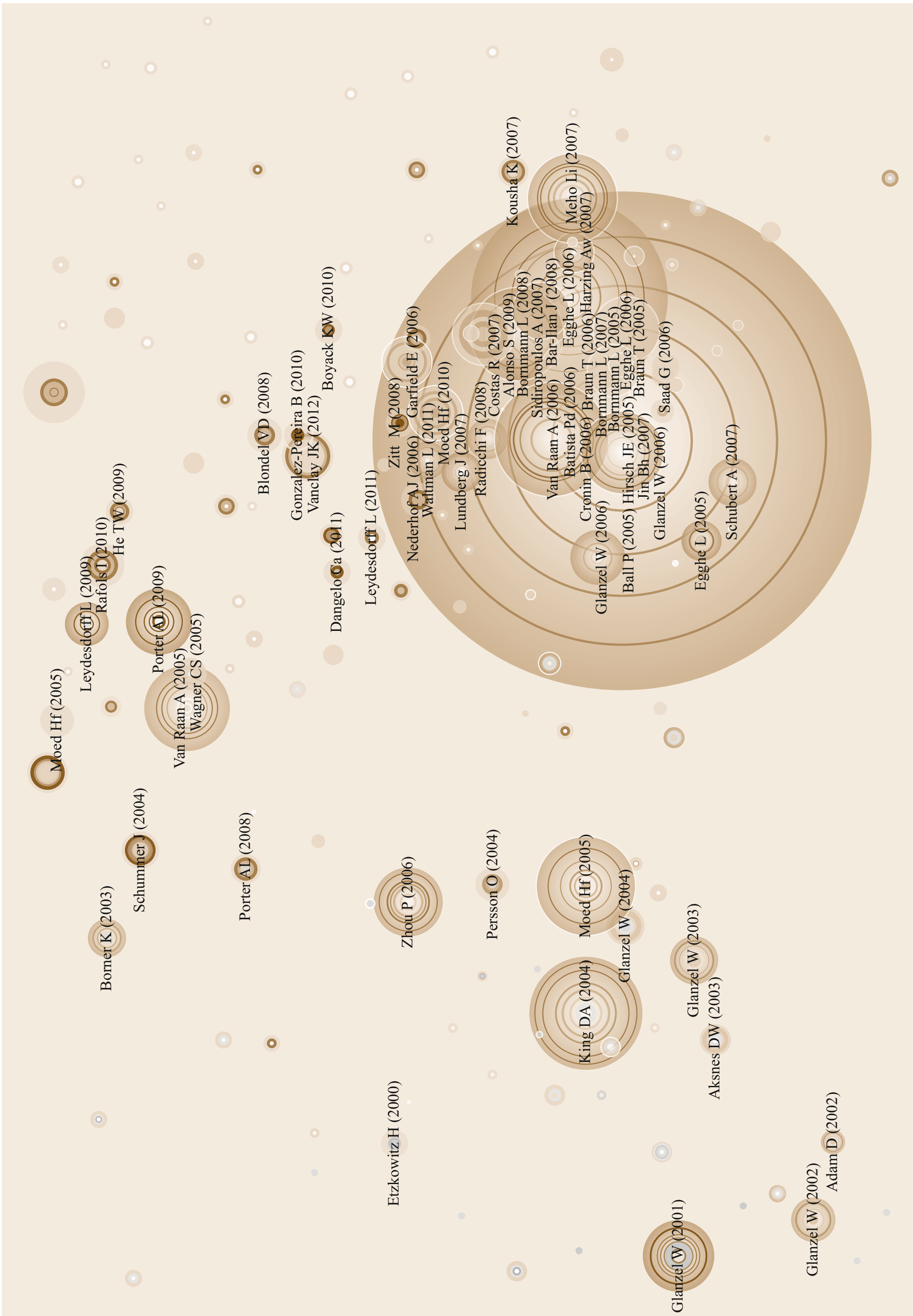


Fig. 7.3 CiteSpace visualization example

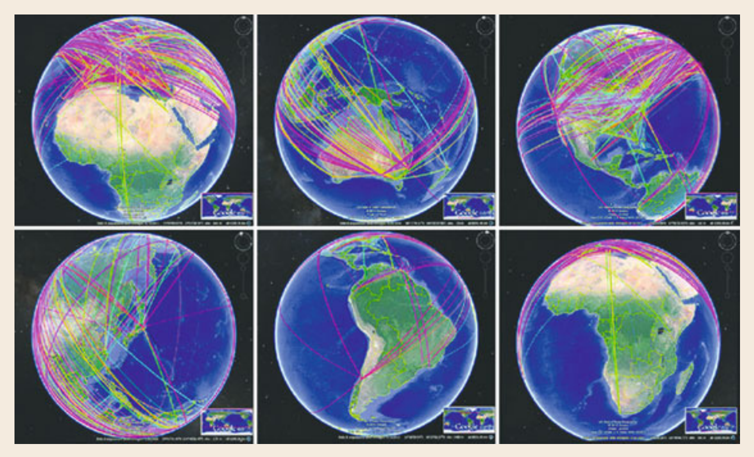


Fig. 7.4 CiteSpace visualization in Google Earth (from [7.69])

CitNetExplorer does not use the title data, because in many cases is not written in a consistent way [7.38], which makes a perfect match between citing and the cited document impossible.

Analysis Options

CitNetExplorer presents three remarkable functionalities:

1. The capability of select publications
2. Drill-down and expand functionalities
3. Different algorithms to generate the network.

When the network obtained contains many citation relations, CitNetExplorer offers the option of using *transitive reduction* to display only a selection. This selection consists in removing all non-essential citation relations from the network. In other words, "a citation relation from publication A to publication B is considered non-essential if other paths from publication A to publication B exist" [7.38]. Although the number of citation relations is lower, it ensures that any pair of publications is still connected in any path.

On the other hand, the *drill-down* function allows us to identify a specific topic inside the research area analyzed. Using this function, only the selected publications will be displayed. This functionality is divided into two steps: publication selection from the current network and network update (including only the documents selected). The network obtained after drilling down is called the current network. Furthermore, there are three different approaches to select documents:

- Select all the documents in a defined time period.
- Select the publications assigned to a group.
- Mark one or more documents in the current network to obtain a new network (default).

Another important functionality is the *expansion*, it could be considered as the opposite to *drill-down*. It consists on add publications to the current network that are closely linked to the publications that compose the network. It is useful when the user wants to expand the subnetwork (current network after drill-down) to include publications that are not directly about the topic but are related to it.

In relation to these functionalities, several consecutive drill-downs and expansions could be performed using CitNetExplorer. It allows to obtain different subnetworks composed by the documents that form an specific topic or subtopic.

In relation to the analysis options offered by CitNetExplore, four different options are provided:

1. Extract connected components
2. Cluster publications
3. Identify core publications
4. Find the shortest or the longest path from a publication to another.

Through the clustering process, each publication that is comprised in the network is classified in a cluster, according to the closeness among the different items. Thus, clusters are composed by publications strongly connected by means of their citation relations. Each cluster is then considered as a topic in the research area studied. The publications are assigned by *maximizing* [7.76].

Another analysis option is the identification of core publications. By means of the concept of *k-cores* [7.77], CitNetExplorer identifies publications that compose the core of a network. This allows us to disregard unimportant publications in the periphery of the network obtained. The user can establish the minimum number of citation relations (incoming or outcoming) that a publication should have to be considered as a core publication.

Visualization

The visualization interface allows us to interact with the generated network in different ways. These interactions are useful in order to explore and obtain a deeper insight of the information offered. It allows us to explore the network units using the zoom and scroll functionality and incorporates a smart labeling algorithm to avoid overlapping; if the mouse cursor is put over a node, the citation subnetwork is highlighted. On the other hand, not only the direct citation relations are visible, but the visualization can be set to show higher-order indirect citation relations.

In order to create the network visualization, each document is represented in horizontal and vertical dimensions (Fig. 7.5). To do this kind of visualization, CitNetExplorer follows the research literature on hierarchical graph drawing [7.78]. Documents are placed in vertical dimension based on the publication year. All documents published in the same year are assigned to the same layer, unless the number of publications exceeds the maximum (10 by default). This process is accomplished using a simple heuristic algorithm that minimizes the number of layers. On the other hand, documents are placed in horizontal dimension based on their closeness to each other in the network. The horizontal location is determined by *minimizing* [7.79].

7.3.4 SciMAT

SciMAT (science mapping analysis software tool) is an open source (General Public License (GPLv3)) science mapping software tool designed to assist all the steps in the science mapping workflow, incorporating all the necessary elements (methods, algorithms, and measures) to obtain the different analyses and visualizations. SciMAT was developed by the SECABA Lab at University of Granada, Spain.

SciMAT is based on a longitudinal science mapping approach that establishes the following four steps [7.10]:

1. Research theme detection
2. Low-dimensional space layout of research themes
3. Discovery of thematic areas
4. Performance analysis.

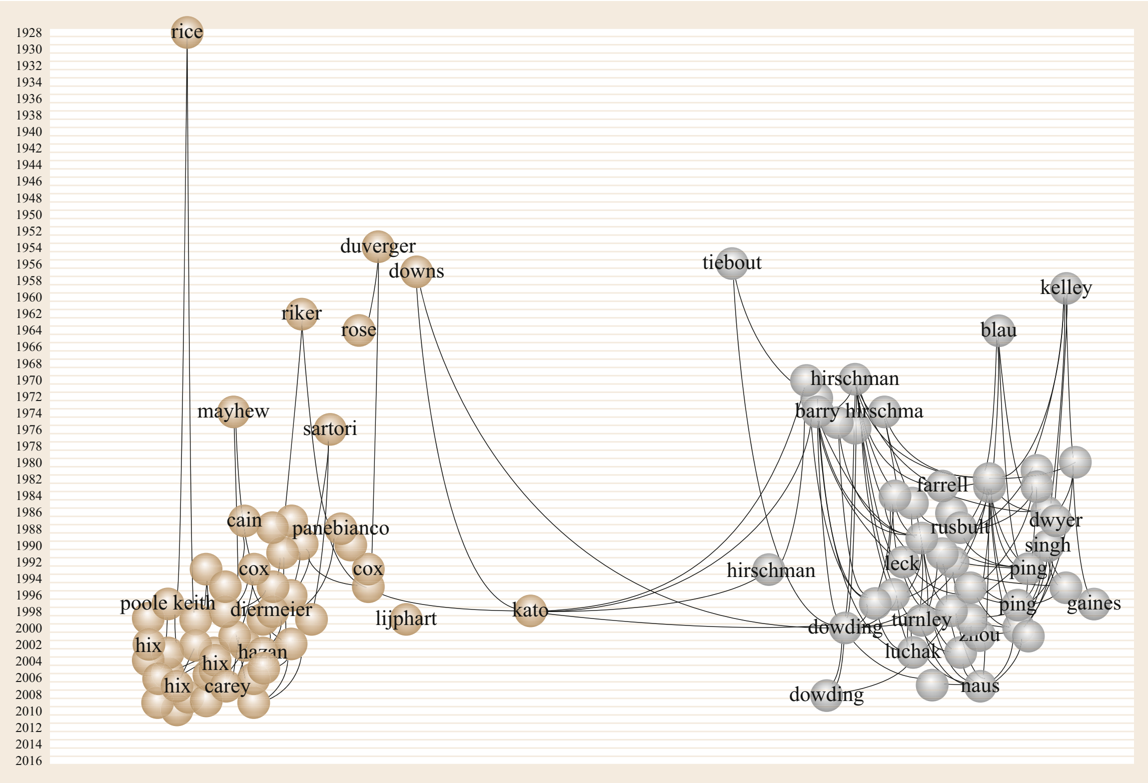


Fig. 7.5 CiteNetExplorer visualization example

Moreover, it has some remarkable characteristics:

- It incorporates all modules to perform all the steps of the science mapping workflow. It supports the analyst in carrying out all the different steps, from the data loading to the visualization and interpretation of the output.
- It incorporates the majority of methods to build bibliometric networks, different similarity measures, and several visualization techniques.
- It allows us to perform different preprocessing tasks, such as detecting duplicate and misspelled items, time slicing, and data and network reduction.
- The analysis is performed in a longitudinal framework in order to analyze and track the conceptual, intellectual, or social evolution of a research field through consecutive time periods.
- The science maps obtained are enriched with bibliometric measures based on citations, such as the sum, maximum, minimum, and average citations. Moreover, it uses advanced bibliometric indexes such as the *H*-index [7.11, 80], *G*-index [7.81], HG-index [7.12], and q^2-index [7.82].

As can be found in bibliographic databases, SciMAT has been used to perform different applied analyses in several thematic areas. Some of the recent publications were focused on:

1. Social work [7.83, 84]
2. Intelligent techniques in health systems [7.85]
3. Virtual and remote labs in education [7.86]
4. Software product lines [7.86]
5. Animal science [7.87]
6. The *Knowledge-Based Systems* journal [7.88]
7. Qualitative marketing [7.89]
8. Integrative and complementary medicine [7.90]
9. Intelligent transportation systems [7.91, 92].

In view of the many applied studies, it can be considered as a useful tool to perform science mapping analyses.

Data Processing

SciMAT allows users to add files in WoS and RIS (research information systems) formats. It then incorporates a preprocessing module where de-duplicating (manual, by plural or by Levenshtein distance, or importing from a XML (extensible markup language) file), time-slicing, data reduction, and network reduction can be performed.

When the files are added to SciMAT, it generates a knowledge base composed of 16 entities:

1. Affiliation
2. Author
3. Author group
4. Author reference
5. Author reference group
6. Document
7. Journal
8. Publication date
9. Period
10. Reference
11. Reference group
12. Reference source
13. Reference source group
14. Subject category
15. Word
16. Word group.

From these entities, five can be used as a unit of analysis in the science mapping analysis carried out by SciMAT:

1. Author
2. Word
3. Reference
4. Author reference
5. Source reference.

Thus, the entities involved in the analysis should be carefully preprocessed, paying attention to the misspelling and de-duplicating process. The de-duplicating process joins the similar items, so only one of them remains. When two items are joined, only one of them is kept in the knowledge base, and it is impossible to know the initial item joined. For this reason, SciMAT incorporates the concept of group for each unit of analysis. A group is a set of items that represents the same entity.

Analysis Options

According to the literature published on SciMAT [7.39], it divides the analysis process into four main stages:

1. *Build the dataset*. The user configures the periods of time used in the analysis, the aspects to analyze (author group, author reference group, source reference group, reference group, or word group) and the portion of data used (minimum frequency as a threshold).
2. *Create and normalize the network*. The network is built using co-occurrence or coupling relations or, indeed, aggregating coupling. Then, the network is filtered to keep only the most representative items.

Finally, a normalization process is performed using a similarity measure.

3. *Apply a clustering algorithm to obtain the map and its associated clusters or subnetworks*. Different cluster algorithms to build the map are available in SciMAT, such as a simple center algorithm [7.7, 17], single linkage [7.93], and variants such as complete linkage, average linkage, and sum linkage.
4. *Apply a set of analyses*. In this last step, different analyses can be performed on the generated map (network, performance, and longitudinal analyses).

Visualization

Different visualization techniques are available in SciMAT, such as strategic diagrams, cluster networks, evolution maps, and overlapping maps. In the following paragraphs, some characteristics of the different visualization techniques are described.

The strategic diagram represents the detected clusters of each period in two-dimensional (2-D) space and categorizes them according to their Callon density and centrality measures. Each cluster shown in the strategic diagram can be enriched by the previously selected bibliometric data measured (Fig. 7.6, extracted from *Moral-Munoz* et al. [7.90]). Furthermore, the associated network for each cluster is also obtained (Fig. 7.7).

Moreover, the results of the temporal or longitudinal analyses are shown using an evolution map and an overlapping-items graph. In Fig. 7.8, the evolution map obtained by *Moral-Munoz* et al. [7.90] is shown.

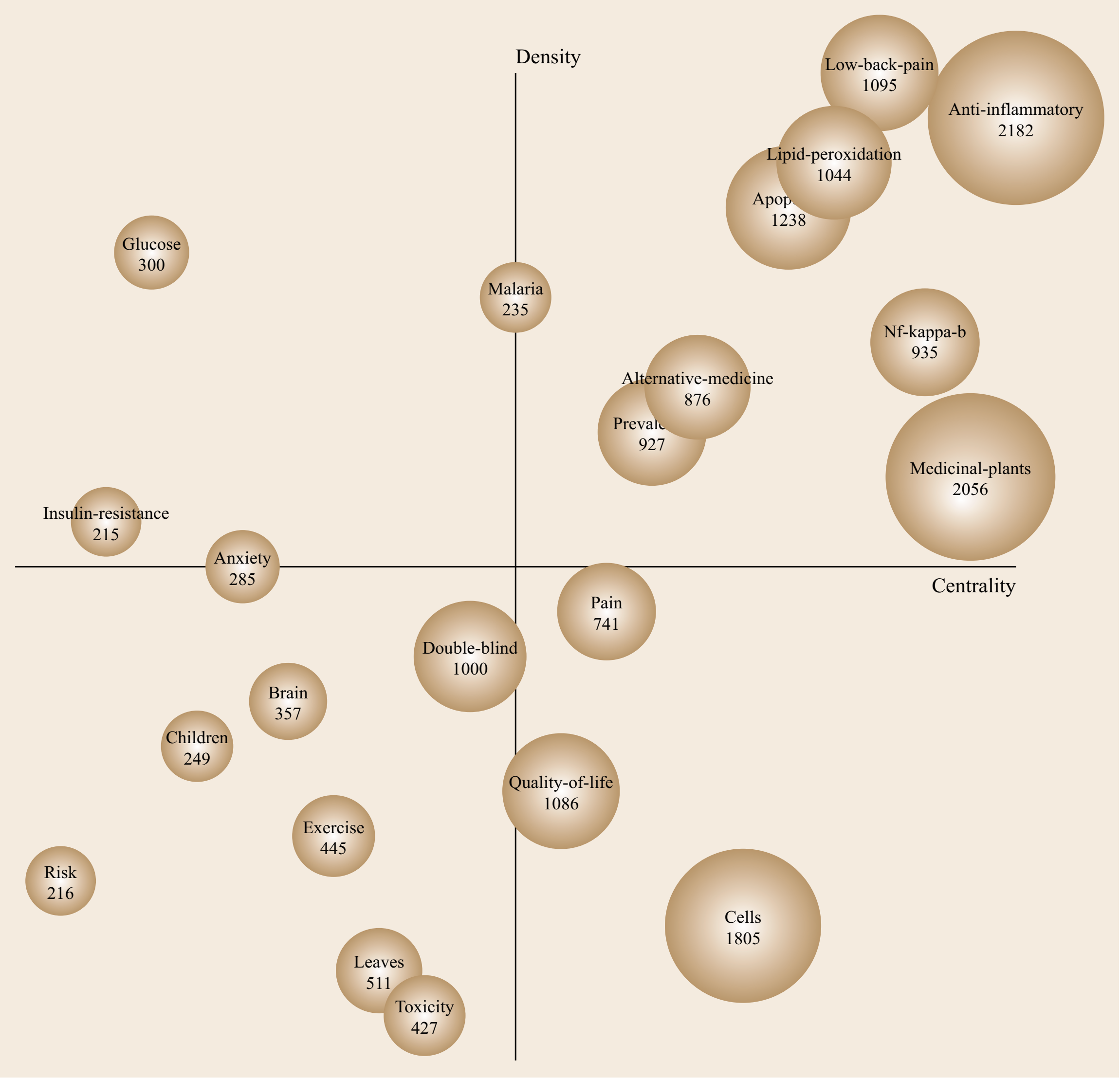


Fig. 7.6 SciMAT strategic diagram [7.90]

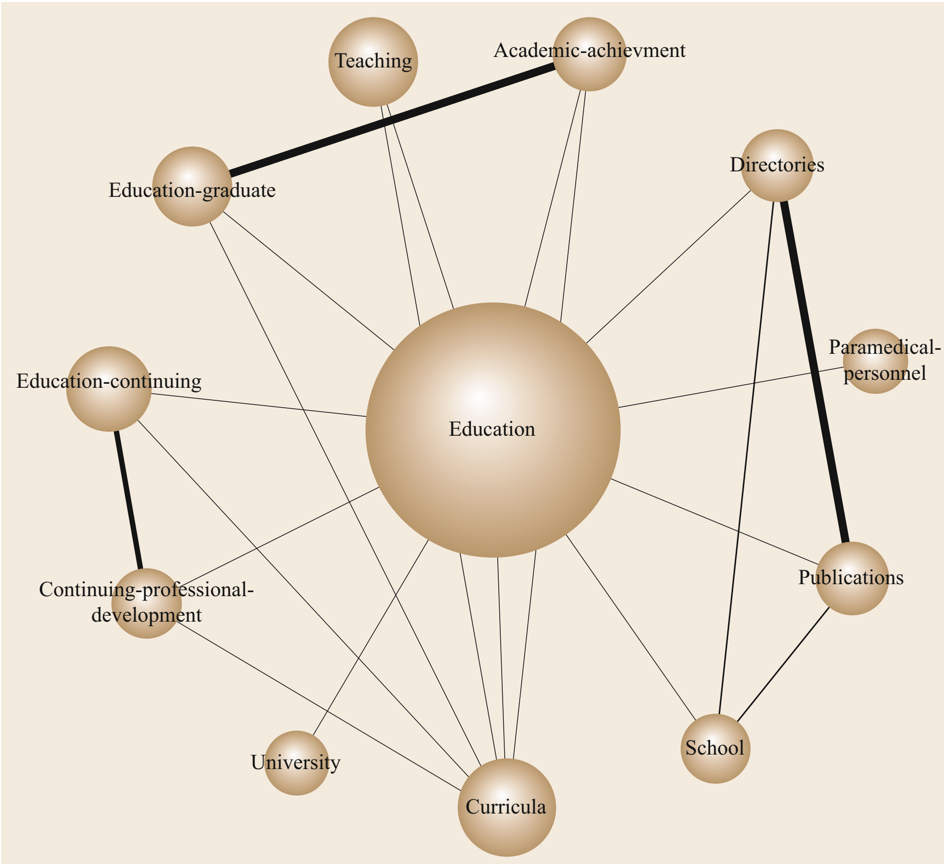


Fig. 7.7 SciMAT cluster network

Furthermore, in Fig. 7.9, the overlapping-items graph across the three consecutive periods is shown. It represents the periods and their number of associated items. Arrows represent the number of items shared by both periods, the number of new items, and the items that come out of the following period.

Finally, it is interesting to remark that the visualization module can build a report in HTML (hypertext markup language) or LaTeX format. The images (strategic diagrams, overlapping-items map, etc.) are exported in PNG (portable network graphics) and SVG (scalable vector graphics) formats, so the user can edit them easily. Furthermore, the cluster networks and evolution maps are exported in Pajek format.

7.3.5 Sci^2 Tool

Sci^2 Tool [7.40] is a modular toolset that is particularly intended for the research of science. It supports temporal, geospatial, topical, and network analysis and the representation of datasets at the micro (individual), meso (local), and macro (global) levels. It was developed by the Cyberinfrastructure for Network Science Center at Indiana University, USA.

Sci^2 Tool specifically focused on scientific documents, and it incorporates algorithms to treat this type of analyses. The main characteristic of this tool is the modular configuration, which allows preparation of the data to submit it to a posterior analysis. Furthermore, some plugins can be added to perform different tasks.

Several applied studies have been published using the Science of Science (Sci^2) tool. Some recent publications have focused on the following thematic areas:

1. Social innovation research [7.94]
2. Twitter conversations [7.95]
3. Engineering education [7.96]
4. Global positioning system [7.97]
5. Innovation systems [7.98]
6. Scholarly publications [7.99]
7. Project management [7.100].

In view of this production, it worth considering it an adequate tool to perform science mapping analyses.

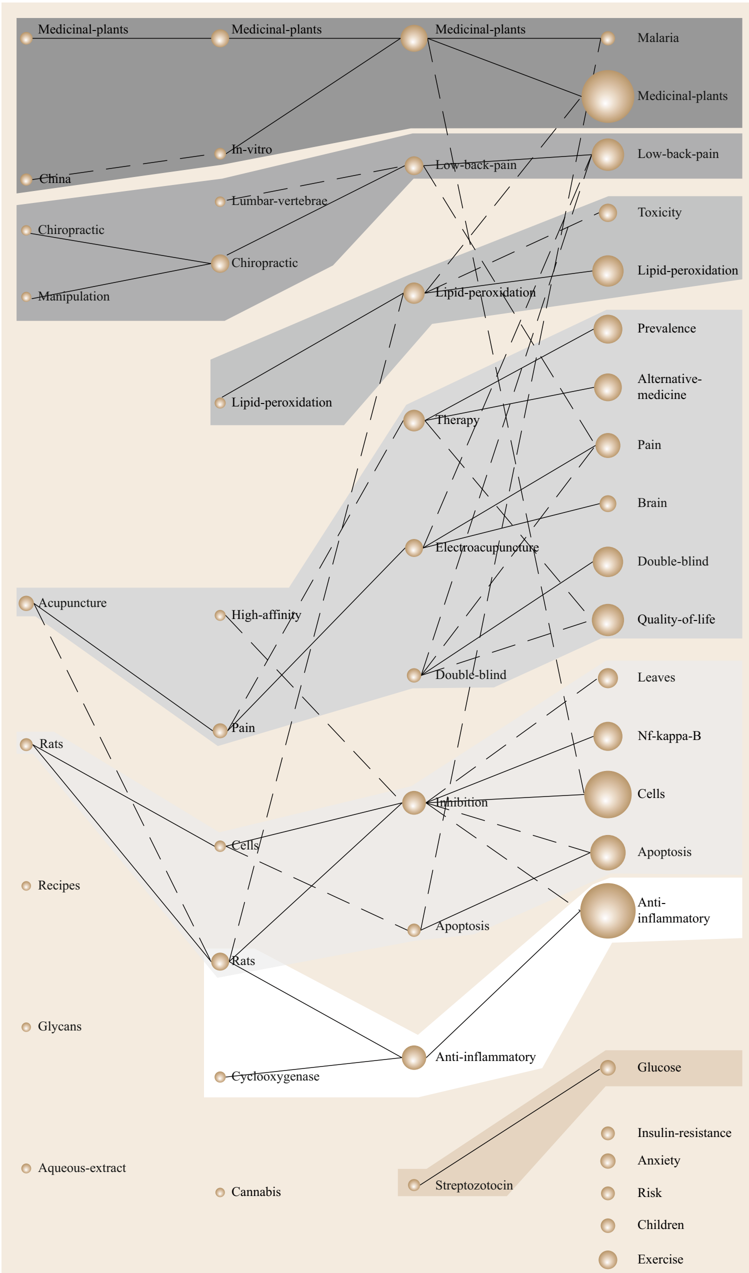


Fig. 7.8 SciMAT evolution map example [7.90]

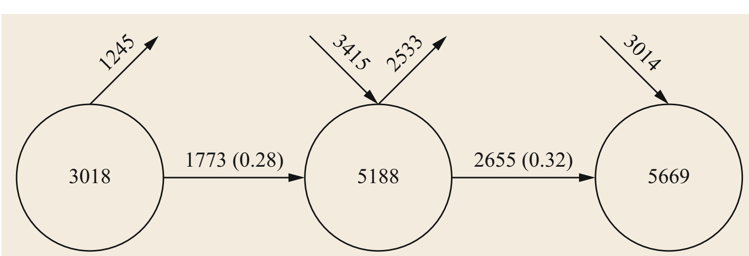


Fig. 7.9 SciMAT overlapping map

Data Processing

Sci^2 Tool is able to read several bibliographic data formats, such as WoS, Scopus, Google Scholar, Bibtex, and the exportation data format of EndNote. Furthermore, it can analyze data information from social media such as Facebook, research funding from the NSF and the National Institutes of Health, as well as other academic data in CSV (comma-separated values) format. As can be observed, this tool supports a wide variety of information sources.

Once the data has been added to this tool, it offers the possibility to submit this data to several preprocessing tasks. They depend on the type of analysis the user wants to perform:

1. In temporal analysis, a slice by time can be performed.
2. In geospatial analysis, the ZIP code can be extracted.
3. In topical analysis, different words and transformations are allowed (lowercase, tokenize, stem, and stopwords).
4. In network analysis, there are several options, such as to extract top nodes, delete isolates, or apply the Pathfinder algorithm.

Analysis Options

The Sci^2 Tool workflow is based on the typical science study [7.13, 40]:

1. Data acquisition and processing
2. Data analysis
3. Modeling
4. Layout.

Finally, the results obtained need to be checked in collaboration with domain experts.

This tool allows extraction of different types of networks and performing several analysis (temporal, geospatial, textual, and networks). Mainly, Sci^2 Tool obtains the following bibliometric networks:

1. Co-author
2. Co-PI (principal investigator)
3. Document co-citation
4. Journal co-citation
5. Author co-citation
6. Bibliographic coupling
7. Author bibliographic coupling
8. Journal bibliographic coupling.

Likewise, this tool allows building direct link networks, such as author references, document references, journal references, and author document networks.

Visualization

In order to represent the networks obtained, Sci^2 Tool applies several algorithms to map and analyze them. As stated above, there are four types of analysis, and consequently, different visualizations can be obtained.

Some characteristics in relation to the different visualizations are briefly described below:

1. *Temporal visualization* (Fig. 7.10, extracted from Sci^2 Tool documentation [7.40]). Through data slicing, the user can evaluate the evolution of the network. The result is presented as a horizontal bar graph. In this type of visualization, the *x*-axis is time, and the *y*-axis is amount. Moreover, the visualization output consists of bars with labels and colors than correspond to the different units of analysis.
2. *Geospatial visualization*. Using the ZIP code extraction, the user can build a geospatial map where an author's location is presented. Sci^2 Tool offers two geospatial options: the choropleth map and the proportional symbol map:

   - The choropleth map (Fig. 7.11): This presents color countries of the world or states of the US in proportion to numerical data. The user can also scale each individual dimension logarithmically or exponentially and add legends for each visualization dimension.
   - Proportional symbol map: This shows the geospatial information associated to three numerical attributes, which are visualized as symbols overlaid on a world or US base map. The size and colors are proportional to the attributes' values. As in the above case, each individual dimension can be scaled logarithmically or exponentially, and a legend for each dimension can be added.

3. *Topical visualization* (Fig. 7.12). Through the word co-occurrence network, a representation about how the concepts or terms are related is obtained. Then, the map of science represents the subdiscipline nodes that are aggregated to different main disciplines of the area studied. Each discipline has a different color and is labeled. Circles are proportional to the number of records. The lines represent the relations among nodes.
4. *Network visualization* (Fig. 7.13). In this kind of visualization, authors, institutions, and countries, as well as words, papers, journals, patents, and funding are represented as nodes, and their complex relationships as edges. Sci^2 Tool uses the GUESS (the Graph Exploration system) window to show the visualization, and the element attributes of the network can be modified.

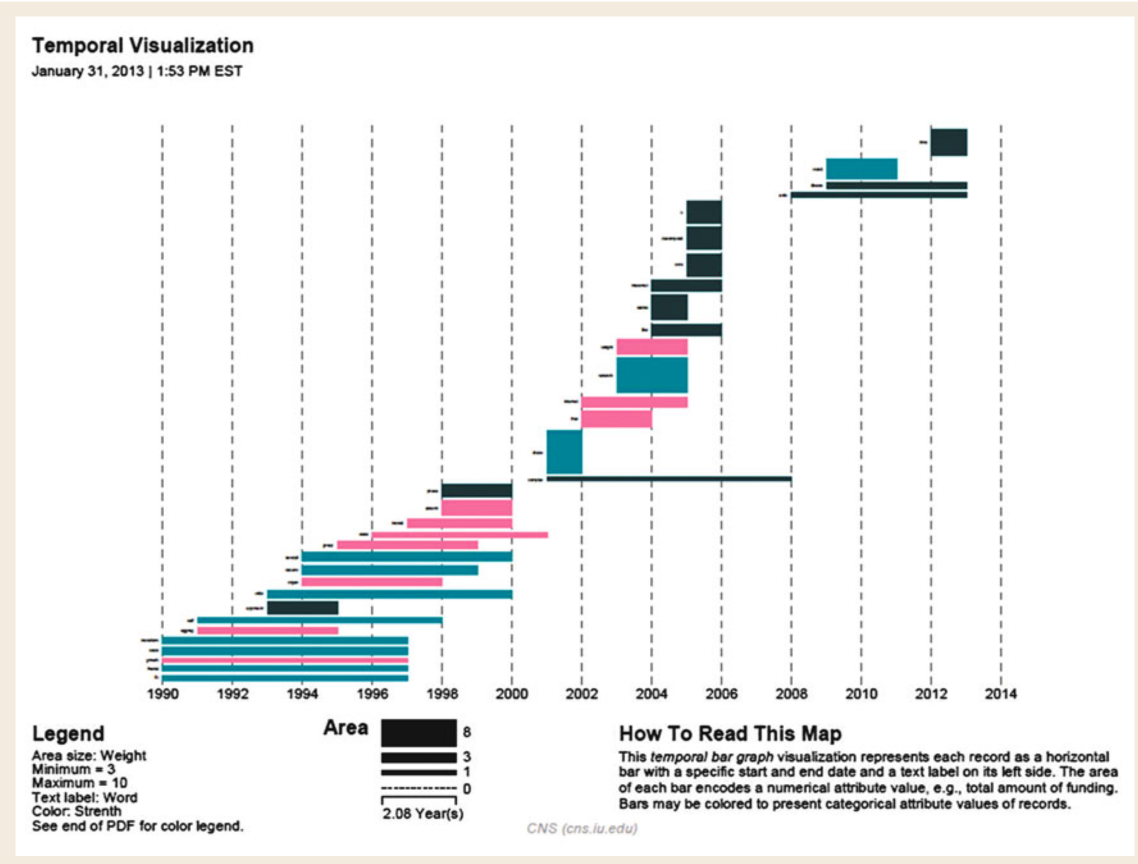


Fig. 7.10 The Sci^2 Tool: temporal visualization [7.40]

In light of the different visualization options presented, it is evident that this tool offers a great and adequate number of possibilities to represent the different aspects of science. It covers the main network visualizations present in the scientific literature.

7.3.6 VOSviewer

VOSviewer [7.38] is a software tool designed for constructing and visualizing bibliometric networks, with journals, researchers, or individual publications as actors, and based on co-citation, bibliographic coupling, or co-authorship relations. It also offers the possibility of building co-occurrence networks of important terms extracted from a corpus of scientific literature, using a text mining functionality. It was developed by the CWTS at Leiden University, The Netherlands.

It is important to highlight that VOSviewer specially focuses attention on the graphical representation of bibliometric maps. Thus, it is helpful for interpretation of a large bibliometric network in an easy way by means of the different visualization options. It allows use of different functionalities, such as zooming, scrolling, and searching.

VOSviewer counts with a large path in the field of science analysis. With the aim of showing the different applications of this software to specific research areas, a search has been conducted. Only in 2016, VOSviewer was used to analyze:

1. The concept of international competitiveness [7.102]
2. Thermal spray [7.103]
3. Triple negative breast cancer [7.104]
4. Antioxidative herbal medicines in type 2 diabetes mellitus [7.105]
5. Space research [7.44]
6. Autism spectrum disorders [7.106]
7. Probiotics in pediatrics [7.107]
8. Spinal tuberculosis [7.108]
9. Iranian papers on reproductive medicine [7.109]
10. Zebrafish in Brazilian science [7.110]
11. Production on open access [7.46]
12. Evidence-based antioxidative herbal medicines in the management of diabetic nephropathy [7.111]

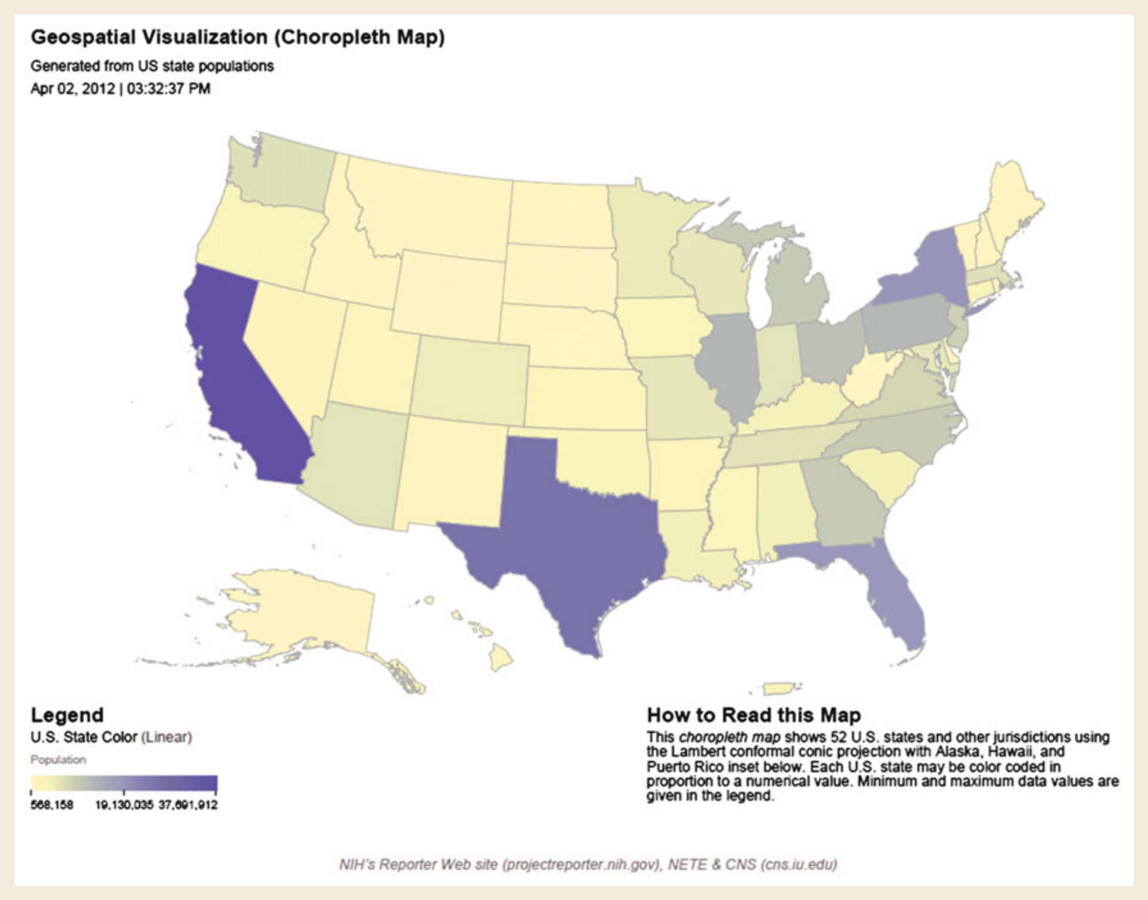


Fig. 7.11 Sci^2 Tool: geospatial visualization [7.40]

13. Stem cells [7.48]
14. Safety science [7.112]
15. Primary care databases [7.113]
16. Green development models [7.114].

As can be observed, it is a software with a large number of studies published. Thus, it is one of the best options for performing a science mapping analysis.

Data Processing

VOSviewer [7.41] can extract bibliographic networks (co-authorship, co-occurrence, and citation-based ones) from bibliographic data. This data is added from files downloaded from WoS, Scopus, PubMed, and in RIS format. Furthermore, it is able to import and export network data from GML (Graph Modeling Language) and Pajek formats.

Analysis Options

VOSviewer incorporates advanced layout and clustering techniques which can be set up using several parameters. According to the literature [7.41], VOSviewer constructs the map based on a co-occurrence matrix in three steps:

1. *Similarity matrix*. In order to apply the VOS mapping technique [7.115], a similarity matrix is needed as input. VOSviewer uses a similarity measure known as the association strength [7.116]. This measure is sometimes referred as to the proximity index [7.61, 117] or probabilistic affinity index [7.118].
2. *VOS mapping technique*. A two-dimensional map is constructed in which the items are located in such a way that the distance between items reflects the similarity measure approximately, using the VOS mapping technique [7.115]. Therefore, if two items have a higher similarity they should appear closer.
3. *Translation, rotation, and reflection*. In order to correct the optimization problem described in the literature [7.119], the map is submitted to different solutions: translation, rotation and reflection. These transformations are enough to ensure the results consistence.

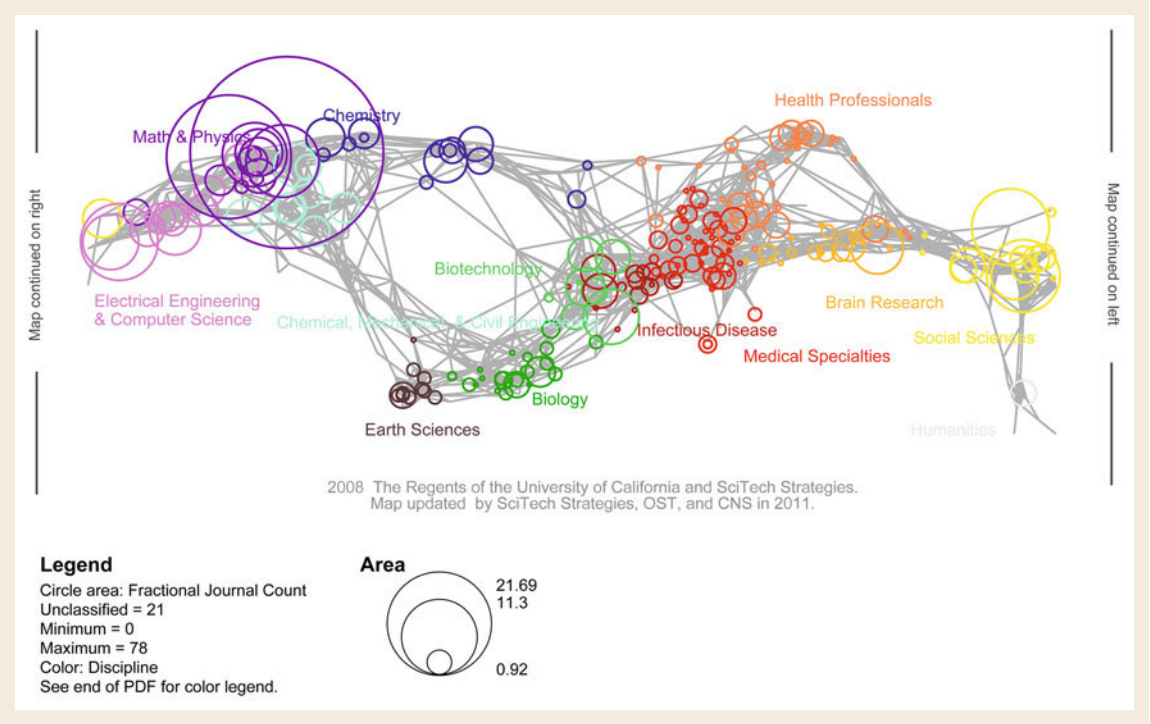


Fig. 7.12 Sci^2 Tool: topical visualization [7.101]

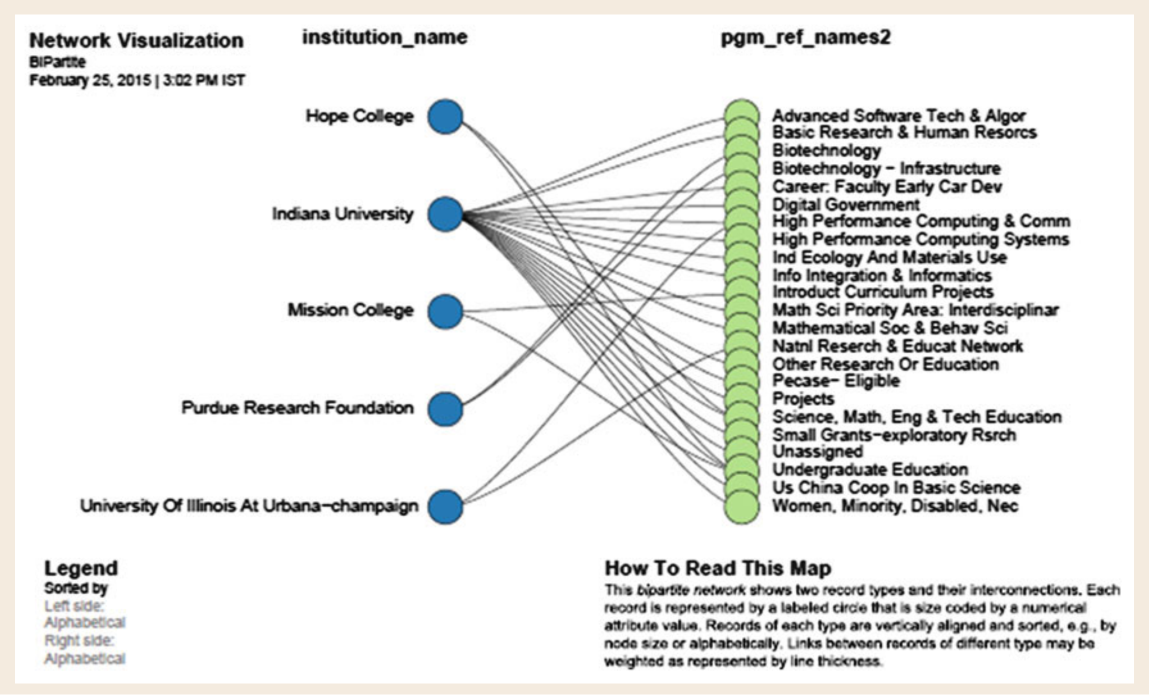


Fig. 7.13 Sci^2 Tool: network visualization [7.40]

Visualization

Distance-based maps were chosen to build the bibliometric maps [7.38]. In these kinds of maps, the distance between two items represents the strength of the relation between them. If an item is close to another, it indicates a stronger relation. Among these types of mapping techniques, there are several options, such as multidimensional scaling [7.119], VxOrd [7.120, 121], and one proposed by *Kopcsa* and *Schiebel* [7.122]. VOSviewer can be employed to view multidimensional scaling maps using statistical software such as SAS (Statistical Analysis System), SPSS, and R. Moreover, the VOS mapping technique [7.115] has been fully integrated, allowing the visualization and construction of this type of map without using other computer program.

In relation to visualization capabilities, this software provides three visualization options:

1. Network
2. Overlay
3. Density.

In what follows, the characteristics of these types of visualizations are described [7.123]:

1. In the *network view* (Fig. 7.14), items are shown by a circle with a label. The volume of the circle and the size of the label depend on the item's importance. Thus, if an item is more important, its label and circle are bigger. The color of the circle is related to the cluster assigned to a group of items. The color is useful for representing some network characteristics such as topic, work group, etc.
2. In the *overlay view* (Fig. 7.15), the network is represented as in the network view, but the items are colored differently. At the bottom part of the main panel, a gradient color bar based on the scores given to the items is shown. These scores can be obtained from the score column in a map file, or the user can specify the items' colors. If any information is given, the overlay view is not available.
3. In the *density view* (Fig. 7.16), there are two different visualization options, which can be selected in the options panel:

   - Item density: Each point in a map has a color determined by the items' density at that point. The map is shown with colors from red to blue. Thus, a larger number of items is represented in red, and a smaller number of items is represented in blue.
   - Cluster density: This is used only if items have been assigned to clusters. An item's density is displayed separately for each cluster. The color is determined by the cluster's color and is more or less intense depending on the weighting.

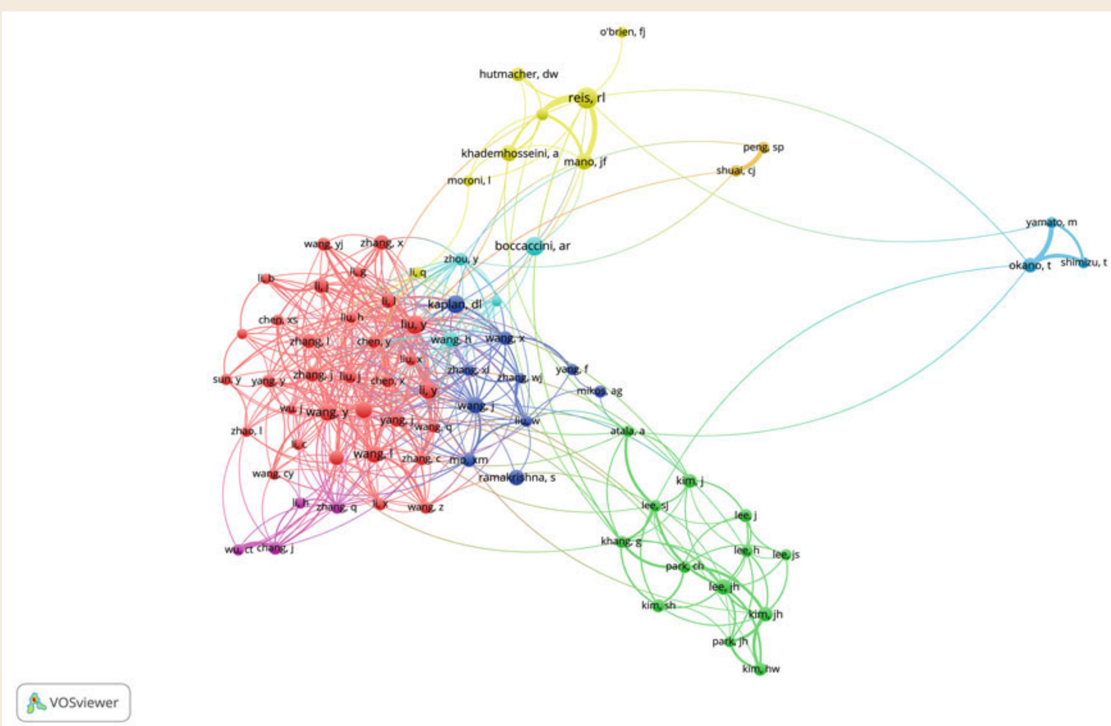


Fig. 7.14 VOSviewer: network view

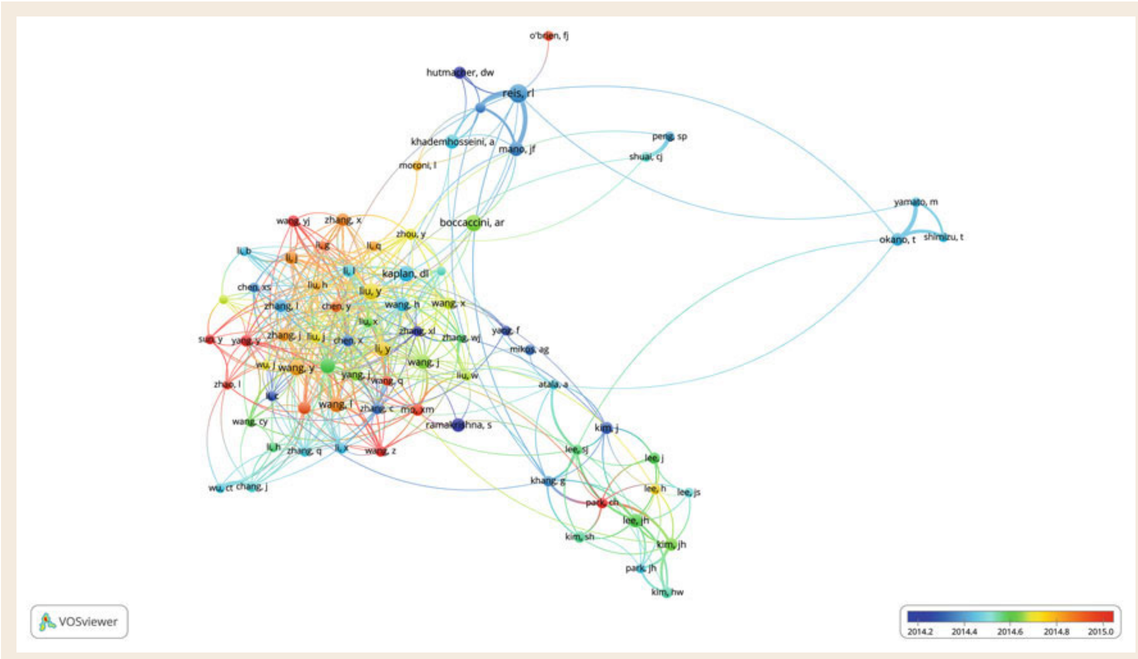


Fig. 7.15 VOSviewer: overlay view

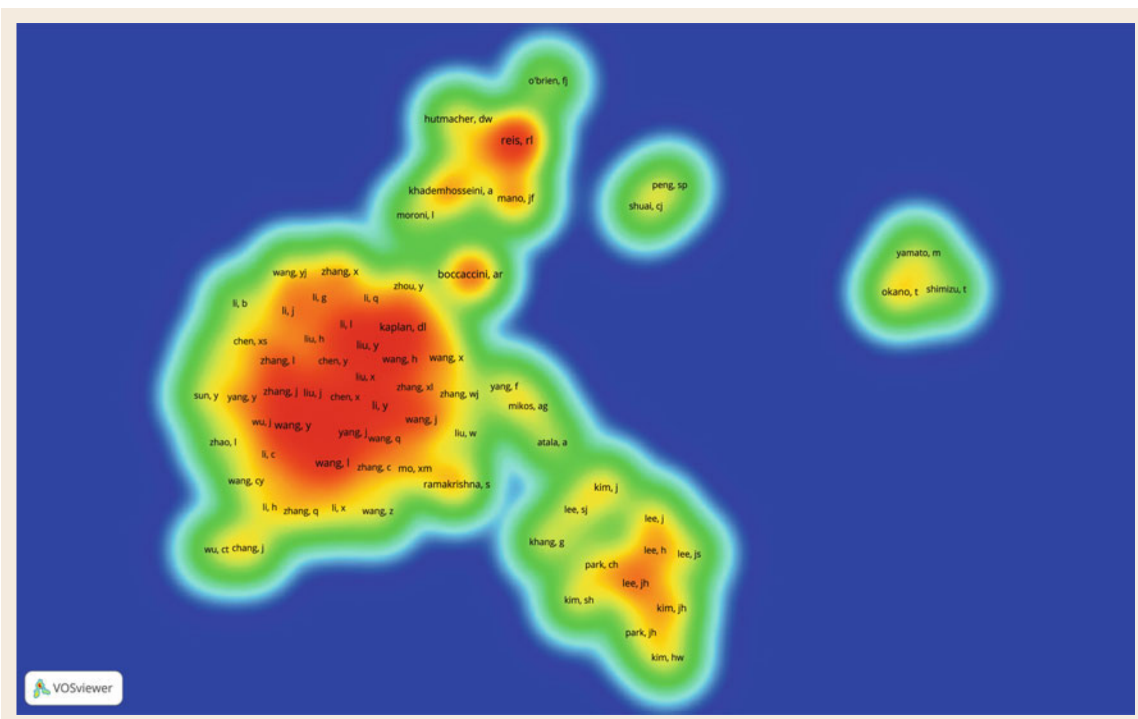


Fig. 7.16 VOSviewer: density view

Moreover, the zoom and scroll functionality and smart labeling algorithm prevent labels from overlapping. VOSviewer incorporates the zoom and scroll option in order to facilitate a detailed examination of the map generated. This functionality used in three ways:

1. Using the mouse
2. Using the navigation buttons (upper right corner)
3. Using the keyboard.

Furthermore, because labels tend to overlap in a large amount of network items, the software incorporates a labeling algorithm to prevent this.

Finally, all of the visualizations generated can be saved in different graphical file formats, such as bitmaps or vectors. This option makes it easy to include the analysis output in any format, digital or printed.

7.4 Software Characteristics: Summary and Comparison

Once the main characteristics of the software have been described, a comparison of tools' features is shown. In the following paragraphs, different features are discussed and evaluated:

1. Operative system
2. Data sources
3. Preprocessing
4. Bibliometric analysis
5. Bibliographic networks
6. Normalization.

According to Table 7.3, all the software run on Windows. Nevertheless, there are only two types of software exclusively for this system, BibExcel and CiteSpace, the rest are multiplatform ones (Windows, OSX, and Linux). On the other hand, they can all use ISIWoS and Scopus as the source. If a user wants to perform an analysis from a different database, a tool that allows to perform the analysis or tries to transform the data file must be selected.

Table 7.3 Operative systems and data source of the science mapping tools

Software tool	Operative system	Data source
BibExcel	Win	ISIWoS, Scopus, proCite, BibExcel format
CiteSpace II	Win	ISIWoS, Scopus, PubMed, arXiv, patents
CitNetExplorer	Win, OSX, Linux	ISIWoS, Scopus
SciMAT	Win, OSX, Linux	ISIWos, Scopus
Sci^2 Tool	Win, OSX, Linux	ISIWoS, Scopus, Google Scholar, Bibtex, EndNote
VOSviewer	Win, OSX, Linux	ISIWoS, Scopus, PubMed

In Table 7.4 some characteristics in relation to the analysis options offered by the different tools are shown. Several preprocessing options are available in the tools analyzed. Data and network reduction is available in all of them, but there are other options, such as time slicing (e.g. CiteSpace II, SciMAT, or Sci^2 Tool) and duplicate detection (e.g. SciMAT and Sci^2 Tool). In relation to bibliographic networks, most of them are able to manage co-citation, coupling, and collaboration networks, only CitNetExplorer is specifically centered on co-citation networks. Finally, it is worth highlighting that there are several possibilities with regard to the normalization process. The two software from the CWTS, CitNetExplorer, and VOSviewer use the association strength as a normalization measure. Salton's cosine and Jaccard's index is available in BibExcel, CiteSpace II, and SciMAT. Furthermore, there are other, less frequent, measures, such as Vladutz and Cook measures in BibExcel, Dice index in CiteSpace II, or the equivalence index in SciMAT.

It is well known that science mapping analysis is characterized by the visualization of the data. From Table 7.5, it can be observed that the visualization output in each software is totally different. Each software uses a kind of visualization in order to represent the information in an adequate manner. There are several options, the most common being network visualization, but other interesting ones are tree rings, geospatial maps, or evolution maps. These visualizations will fit to the type of actor to be analyzed.

Table 7.4 Analysis options of the science mapping tools

Software tool	Preprocessing	Bibliographic network	Normalization
BibExcel	Data and network reduction	Co-citation, coupling, collaboration	Salton's cosine, Jaccard's index, or Vladutz and Cook measures
CiteSpace II	Time slicing and data and networks reduction	Co-citation, coupling, collaboration	Salton's cosine, Dice or Jaccard strength
CitNetExplorer	Data and network reduction	Co-citation	Association strength
SciMAT	Duplicate and misspelled items detection, time slicing, data and network reduction	Co-citation, coupling, collaboration	Association strength, equivalence index, inclusion index, Jaccard index, and Salton's cosine
Sci^2 Tool	Duplication detection, time slicing, and data and networks reduction	Co-citation, coupling, collaboration	User defined
VOSviewer	Data and network reduction	Co-citation, coupling, collaboration	Association strength

Table 7.5 Visualization options of the science mapping tools

Software tool	Visualizations
BibExcel	**External software*
CiteSpace II	Tree ring, geospatial map
CitNetExplorer	Network
SciMAT	Strategic diagram, cluster network, overlapping map, evolution map
Sci^2 Tool	Temporal, geospatial map, topical, network
VOSviewer	Network, overlay, density

Therefore, we can state that the variability in measures and network analyses is high. The user needs to know the main characteristics of the software in order to adapt the expectations to the final output. The tools shown in the present review are powerful and offer several possibilities. Each software allows us to analyze and discover different aspects about the network. Thus, it could be useful to use several types of software in order to obtain complementary outputs.

7.5 Conclusions

Bibliometrics is the application of quantitative analysis to publications. This evaluation is used in almost all science fields to measure growth, maturity, the main actors, and conceptual and intellectual development, discovering the trends of a scientific community. In order to perform this kind of analysis, different bibliometric software are needed.

In the present chapter, six tools have been analyzed:

1. BibExcel [7.36]
2. CiteSpace II [7.37]
3. CitNetExplorer [7.38]
4. SciMAT [7.39]
5. Sci^2 Tool [7.40]
6. VOSviewer [7.41].

All of these offer different possibilities to build science maps. There are several differences among data processes, analysis options, and visualizations. The use of one or the other depends on the preferred output and consequent analysis. Although there are software with a wide number of options, none of them implement all the desirable characteristics. There are software that are more focused on visualization and others on analysis. However, they can be used in combination to obtain complementary outputs.

In relation to the data sources, probably CiteSpace II and Sci^2 Tool are the software with the larger number of possibilities. All of the tools offer the option of reducing data and networks, although some of them are able to perform time slicing, such as CiteSpace II, SciMAT, and Sci^2 Tool, and duplicate detection, such as SciMat and Sci^2 Tool. In relation to bibliographic networks, only CitNetExplorer is centered on co-citation; coupling and collaboration networks are also available in the other software. Finally, the normalization measures are varied, and each software offers those that are considered more suitable.

The visualizations offered by the software could cover the user's necessities depending on the expected output. CiteSpace II or Sci^2 Tool are good options to represent a collaboration network; they offer the geospatial map visualization that facilitates the analysis of the publication relationships. If the desired output is to represent temporal development or evolution, SciMAT, Sci^2 Tool, and VOSviewer can be selected, because they incorporate a specific representation. In general terms, all of the software offer an adequate representation of citation networks.

To sum up, there are several freely available science mapping software with a strong potential. They offer all capacities and features to perform this type of analysis during the whole workflow in a more or less easy way. The choice depends on the type of actor to be analyzed and the output expected.

Acknowledgments. The authors would like to acknowledge FEDER funds under grants TIN2013-40658-P and TIN2016-75850-R and also the financial support from the University of Cádiz, Project PR2016-067.

References

7.1 D. Price: *Little Science, Big Science* (Univ. Press, New York, Columbia 1963)

7.2 M.A. Martinez, M.J. Cobo, M. Herrera, E. Herrera-Viedma: Analyzing the scientific evolution of social work using science mapping, Res. Soc. Work Pract. **25**(2), 257–277 (2015)

7.3 J.A. Moral-Munoz, M.J. Cobo, F. Chiclana, A. Collop, E. Herrera-Viedma: Analyzing highly cited papers in intelligent transportation systems, IEEE Trans. Intell. Transp. Syst. **17**(4), 993–1001 (2016)

7.4 H.F. Moed, R.E. De Bruin, T.N. van Leeuwen: New bibliometric tools for the assessment of national research performance: Database description, overview of indicators and first applications, Scientometrics **33**(3), 381–422 (1995)

7.5 M.A. Martinez, M. Herrera, J. López-Gijón, E. Herrera-Viedma: H-Classics: Characterizing the concept of citation classics through H-index, Scientometrics **98**(3), 1971–1983 (2014)

7.6 M. Callon, J.P. Courtial, F. Laville: Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry, Scientometrics **22**(1), 155–205 (1991)

7.7 N. Coulter, I. Monarch, S. Konda: Software engineering as seen through its research literature: A study in co-word analysis, J. Am. Soc. Inf. Sci. **49**(13), 1206–1223 (1998)

7.8 M. Henderson, S. Shurville, K. Fernstrom: The quantitative crunch: The impact of bibliometric research quality assessment exercises on academic development at small conferences, Campus-Wide Inf. Syst. **26**, 149–167 (2009)

7.9 H.F. Moed: New developments in the use of citation analysis in research evaluation, Arch. Immunol. Ther. Exp. **57**(1), 13–18 (2009)

7.10 M.J. Cobo, A.G. López-Herrera, E. Herrera-Viedma, F. Herrera: An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field, J. Infometr. **5**(1), 146–166 (2011)

7.11 J. Hirsch: An index to quantify an individuals scientific research output, Proc. Natl. Acad. Sci. **102**, 15572–16569 (2005)

7.12 S. Alonso, F.J. Cabrerizo, E. Herrera-Viedma, F. Herrera: Hg-index: A new index to characterize the scientific output of research based on the h- and g-indexes, Scientometrics **82**(2), 391–400 (2010)

7.13 K. Börner, C. Chen, K. Boyack: Visualizing knowledge domains, Annu. Rev. Inf. Sci. Technol. **37**(1), 179–255 (2003)

7.14 S. Morris: B. Van Der Veer Martens: Mapping research specialities, Annu. Rev. Inf. Sci. Technol. **42**(1), 213–295 (2008)

7.15 E.C.M. Noyons, H.F. Moed, M. Luwel: Combining mapping and citation analysis for evaluative bibliometric purposes: A bibliometric study, J. Am. Soc. Inf. Sci. **50**(2), 115–131 (1999)

7.16 H. Small: Update on science mapping: Creating large document spaces, Scientometrics **38**(2), 275–293 (1997)

7.17 M.J. Cobo, A.G. López-Herrera, E. Herrera-Viedma, F. Herrera: Science mapping software tools: Review, analysis, and cooperative study among tools, J. Assoc. Inf. Sci. Technol. **62**(7), 1382–1402 (2011)

7.18 V. Batagelj, A. Mrvar: Pajek–Program for large network analysis, Connections **21**(2), 47–57 (1998)

7.19 M. Bastian, S. Heymann, M. Jacomy: Gephi: An open source software for exploring and manipulating networks. In: *Proc. Int. AAAI Conf. Weblogs Soc. Media* (2009)

7.20 M. Aria, C. Cuccurullo: Bibliometrix, http://www.bibliometrix.org/ (Univ. of Naples Federico II, 2016)

7.21 M. Gagolewski: Bibliometric impact assessment with R and the CITAN package, J. Informetr. **5**(4), 678–692 (2011)

7.22 A. Uddin, J. Bhoosreddy, M.M. Tiwari, V.K. Singh: A sciento-text framework to characterize research strength of institutions at fine-grained thematic area level, Scientometrics **106**(3), 1135–1150 (2016)

7.23 S. Grauwin: BiblioTools, http://www.sebastian-grauwin.com/?page_id=427 (Univ. of Lyon, 2014)

7.24 M.J. Cobo, A.G. López-Herrera, E. Herrera-Viedma: A relational database model for science mapping analysis, Acta Polytech. Hung. **12**(6), 43–62 (2015)

7.25 N. Mallig: A relational database for bibliometric analysis, J. Informetr. **4**(4), 564–580 (2010)

7.26 V. Gómez-Jáuregui, C. Gómez-Jáuregui, C. Manchado, C. Otero: Information management and improvement of citation indices, Int. J. Inf. Manag. **34**(2), 257–271 (2014)

7.27 A.L. Porter, S.W. Cunningham: *Tech Mining: Exploiting New Technologies for Competitive Advantage* (Wiley, Hoboken 2004)

7.28 A. Thor, W. Marx, L. Leydesdorff, L. Bornmann: Introducing cited references explorer (CRExplorer): A program for reference publication year spectroscopy with cited references standardization, J. Informetr. **10**(2), 503–515 (2016)

7.29 L. Leydesdorff, L. Bornmann, W. Marx, S. Milojević: Referenced publication years spectroscopy applied to iMetrics: Scientometrics, Journal of Informetrics,

and a relevant subset of JASIST, J. Informetr. **8**(1), 162–174 (2014)
7.30 N.J. van Eck, L. Waltman: *Visualizing Bibliometric Networks* (Springer, Cham 2014)
7.31 V. Batagelj, M. Cerinšek: On bibliographic networks, Scientometrics **96**(3), 845–864 (2013)
7.32 H. Small: Co-citation in the scientific literature: A new measure of the relationship between two documents, J. Am. Soc. Inf. Sci. **24**(4), 265–269 (1973)
7.33 M.M. Kessler: Bibliographic coupling between scientific papers, Am. Doc. **14**(1), 10–25 (1963)
7.34 W. Gänzel: National characteristics in international scientific co-authorship relations, Scientometrics **51**(1), 69–115 (2001)
7.35 M. Callon, J.P. Courtial, W.A. Turner, S. Bauin: From translation to problematic networks: An introduction to co-word analysis, Soc. Sci. Inf. **22**(2), 191–235 (1983)
7.36 O. Persson, R. Danell, J.W. Schneider: *Celebrating Scholarly Communication Studies: A Festschrift for Olle Persson at his 60th Birthday*, Vol. 5 (International Society for Scientometrics and Informetrics, Leuven 2009) pp. 9–24
7.37 C. Chen: CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature, J. Am. Soc. Inf. Sci. Technol. **57**(3), 359–377 (2006)
7.38 N.J. van Eck, L. Waltman: CitNetExplorer: A new software tool for analyzing and visualizing citation networks, J. Informetr. **8**(4), 802–823 (2014)
7.39 M.J. Cobo, A.G. López-Herrera, E. Herrera-Viedma, F. Herrera: SciMAT: A new science mapping analysis software tool, J. Am. Soc. Inf. Sci. Technol. **3**(8), 1609–1630 (2012)
7.40 Sci2 Team: Science of Science (Sci2) Tool, http://sci.slis.indiana.edu (2009) Indiana University and SciTech Strategies
7.41 N.J. van Eck, L. Waltman: Software survey: VOSviewer, a computer program for bibliometric mapping, Scientometrics **84**(2), 523–553 (2010)
7.42 S.P. Borgatti, M.G. Everett, L.C. Freeman: *UCINET 6 for Windows: Software for social network analysis, http://www.analytictech.com* (Analytic Technologies, Harvard 2002)
7.43 C. Liu, Q. Gui: Mapping intellectual structures and dynamics of transport geography research: A scientometric overview from 1982 to 2014, Scientometrics **109**(1), 159–184 (2016)
7.44 R. Eito-Brun, M.L. Rodriguez: 50 years of space research in Europe: A bibliometric profile of the European Space Agency (ESA), Scientometrics **109**(1), 1–26 (2016)
7.45 F. Quevedo-Silva, E.B. Santos, M.M. Brandão, L. Vils: Estudo Bibliométrico: Orientações sobre sua Aplicação, Rev. Bras. Mark. **15**(02), 246–262 (2016)
7.46 S. Miguel: E. Tannuri de Oliveira, M. Cabrini Grácio: Scientific production on open access: A worldwide bibliometric analysis in the academic and scientific context, Publications **4**(1), 1 (2016)
7.47 X.Y. Zhao, L. Sheng, T.X. Diao, Y. Zhang, L. Wang, Z. Yanjun: Knowledge mapping analysis of ebola research, Bratisl. Med. J. **116**(12), 729–734 (2015)
7.48 R.N. Machado, B. Vargas-Quesada, J. Leta: Intellectual structure in stem cell research: Exploring Brazilian scientific articles from 2001 to 2010, Scientometrics **106**(2), 525–537 (2016)
7.49 G. Salton, M.J. McGill: *Introduction to Modern Information Retrieval* (McGraw-Hill, New York 1983)
7.50 P. Jaccard: Distribution de la Flore Alpine: dans le Bassin des dranses et dans quelques régions voisines, Bull. Soc. Vaud. Sci. Nat. **37**(140), 241–272 (1901)
7.51 G. Vladutz, J. Cook: Bibliographic coupling and subject relatedness, Proc. Am. Soc. Inf. Sci. **21**, 204–207 (1984)
7.52 L. Leydesdorff, O. Persson: Mapping the geography of science: Distribution patterns and networks of relations among cities and institutes, J. Am. Soc. Inf. Sci. Technol. **61**(8), 1622–1634 (2010)
7.53 V. Batagelj, A. Mrvar: *Pajek–Analysis and Visualization of Large Networks* (Springer, Berlin, Heidelberg 2004)
7.54 B. Liu, L. Zhang, X. Wang: Scientometric profile of global rice research during 1985–2014, Curr. Sci. **112**(5), 1003–1011 (2017)
7.55 J. Zhu, W. Hua: Visualizing the knowledge domain of sustainable development research between 1987 and 2015: A bibliometric analysis, Scientometrics **110**(2), 893–914 (2017)
7.56 C. Xiang, Y. Wang, H. Liu: A scientometrics review on nonpoint source pollution research, Ecol. Eng. **99**, 400–408 (2017)
7.57 Q. Zhao, W. Zhang, R. Wang, Y. Wang, D. Ouyang: Research advances in molecular modeling in cyclodextrins, Curr. Pharm. Des. **23**(3), 522–531 (2017)
7.58 X. Li, E. Ma, H. Qu: Knowledge mapping of hospitality research—A visual analysis using CiteSpace, Int. J. Hosp. Manag. **60**, 77–93 (2017)
7.59 R.W. Schvaneveldt, F.T. Durso, D.W. Dearholt: *Network Structures in Proximity Data*, *Psychology of Learning and Motivation* (Academic Press, New York 1989)
7.60 C.M. Morel, S.J. Serruya, G.O. Penna, R. Guimarães: Co-authorship network analysis: A powerful tool for strategic planning of research, development and capacity building programs on neglected diseases, PLoS Negl. Trop. Dis. **3**(8), e501 (2009)
7.61 H.P.F. Peters, A.F.J. van Raan: Co-word-based science maps of chemical engineering. Part I: Representations by direct multidimensional scaling, Res. Policy **22**(1), 23–45 (1993)
7.62 Q. He: Knowledge discovery through co-word analysis, Libr. Trends **48**(1), 133–159 (1999)
7.63 H. Small: A co-citation model of a scientific specialty: A longitudinal study of collagen research, Soc. Stud. Sci. **7**, 139–166 (1977)
7.64 H. Small, M.E.D. Koenig: Journal clustering using a bibliographic coupling method, Inf. Process. Manag. **13**(5), 277–288 (1977)

7.65 B. Jarneving: Bibliographic coupling and its application to research-front and other core documents, J. Informetr. **1**(4), 287–307 (2007)
7.66 L.R. Dice: Measures of the amount of ecologic association between species, Ecology **26**(3), 297–302 (1945)
7.67 J. Kleinberg: Bursty and hierarchical structure in streams, Data Min. Knowl. Discov. **7**(4), 373–397 (2003)
7.68 L.C. Freeman: Centrality in social networks conceptual clarification, Soc. Netw. **1**(3), 215–239 (1978)
7.69 N. Bernabò, L. Greco, M. Mattioli, B. Barboni: A scientometric analysis of reproductive medicine, Scientometrics **109**(1), 103–120 (2016)
7.70 E. Garfield, A.I. Pudovkin, V.S. Istomin: Why do we need algorithmic historiography?, J. Am. Soc. Inf. Sci. Technol. **54**(5), 400–412 (2003)
7.71 F. Wu, R. Li, L. Huang, H. Miao, X. Li: Theme evolution analysis of electrochemical energy storage research based on CitNetExplorer, Scientometrics **110**(1), 113–139 (2017)
7.72 B. Elango: Scientometric analysis of nature nanotechnology, Libr. Hi Tech News **34**(1), 23–30 (2017)
7.73 S. Kusumastuti, M.G.M. Derks, S. Tellier, E. Di Nucci, R. Lund, E.L. Mortensen, R.G.J. Westendorp: Successful ageing: A study of the literature using citation network analysis, Maturitas **93**, 4–12 (2016)
7.74 B. Elango, L. Bornmann, S. Shankar: Study of citation networks in tribology research, Collnet J. Scientometr. Inf. Manag. **10**(1), 71–96 (2016)
7.75 A. Avila-Robinson, N. Islam: Evolution of emerging iPS cell-based therapies for age-related macular degeneration (AMD). In: *2015 Portland Int. Conf. Manag. Eng. Technol. (PICMET)* (2015) pp. 495–510
7.76 L. Waltman, N.J. Van Eck: A smart local moving algorithm for large-scale modularity-based community detection, Eur. Phys. J. B **86**(11), 471 (2013)
7.77 S. Seidman: Network structure and minimum degree, Soc. Netw. **5**(3), 269–287 (1983)
7.78 R. Tamassia: *Handbook of Graph Drawing and Visualization* (CRC, Boca Raton 2014)
7.79 N.J. van Eck, L. Waltman, R. Dekker, J. van den Berg: A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS, J. Assoc. Inf. Sci. Technol. **61**(12), 2405–2416 (2010)
7.80 S. Alonso, F.J. Cabrerizo, E. Herrera-Viedma, F. Herrera: H-index: A review focused in its variants, computation and standardization for different scientific fields, J. Informetr. **3**(4), 273–289 (2009)
7.81 L. Egghe: Theory and practise of the G-index, Scientometrics **69**(1), 131–152 (2006)
7.82 F.J. Cabrerizo, S. Alonso, E. Herrera-Viedma, F. Herrera: q^2-Index: Quantitative and qualitative evaluation based on the number and impact of papers in the Hirsch core, J. Informetr. **4**(1), 23–28 (2010)
7.83 M.A. Martinez, F.L. Rodriguez-Fernández, M.J. Cobo, E. Herrera-Viedma: What is happening in social work according the Web of Science?, Cuad. Trab. Soc. **30**(1), 125–134 (2017)
7.84 M.A. Martinez, M.J. Cobo, M. Herrera, E. Herrera-Viedma: Analyzing the scientific evolution of social work using science mapping, Res. Soc. Work Pract. **25**(2), 257–277 (2015)
7.85 M. Lubicz: Applications of operations research and intelligent techniques in the health systems: a bibliometric analysis. In: *Inf. Syst. Archit. Technol. Proc. 37th Int. Conf. Inf. Syst. Archit. Technol. (ISAT 2016)*, Vol. 523 (Springer, Cham 2017)
7.86 R. Heradio, H. Pérez-Morago, D. Fernández-Amoros, F.J. Cabrerizo, E. Herrera-Viedma: A bibliometric analysis of 20 years of research on software product lines, Inf. Softw. Technol. **72**, 1–15 (2016)
7.87 A. Rodriguez-Ledesma, M.J. Cobo, C. López-Pujalte, E. Herrera-Viedma: An overview of animal science research 1945–2011 through science mapping analysis, J. Animal Breed. Genet. **132**(6), 475–497 (2015)
7.88 M.J. Cobo, M.A. Martinez, M. Gutiérrez-Salcedo, H. Fujita, E. Herrera-Viedma: 25 years at knowledge-based systems: A bibliometric analysis, Knowl. Based Syst. **80**, 3–13 (2015)
7.89 E.M. Murgado-Armenteros, M. Gutiérrez-Salcedo, F.J. Torres-Ruiz, M.J. Cobo: Analysing the conceptual evolution of qualitative marketing research through science mapping analysis, Scientometrics **102**(1), 519–557 (2015)
7.90 J.A. Moral-Muñoz, M.J. Cobo, E. Peis, M. Arroyo-Morales, E. Herrera-Viedma: Analyzing the research in integrative and complementary medicine by means of science mapping, Complement. Ther. Med. **22**(2), 409–418 (2014)
7.91 M.J. Cobo, F. Chiclana, A. Collop, J. Oña, E. Herrera-Viedma: A bibliometric analysis of the intelligent transportation systems research based on science mapping, IEEE Trans. Intell. Transp. Syst. **15**(2), 901–908 (2014)
7.92 M.J. Cobo, A. López-Herrera, F. Herrera, E. Herrera-Viedma: A note on the ITS topicevolution in the period 2000–2009 at T-ITS, IEEE Trans. Intell. Transp. Syst. **13**, 413–420 (2012)
7.93 H. Small, E. Sweeney: Clustering the science citation index using co-citations, Scientometrics **7**(3), 391–409 (1985)
7.94 R.P. van der Have, L. Rubalcaba: Social innovation research: An emerging area of innovation studies?, Res. Policy **45**(9), 1923–1935 (2016)
7.95 G. Nelhans, D.G. Lorentzen: Twitter conversation patterns related to research papers, Inf. Res. **21**(2), 41 (2016)
7.96 S. Nawaz, J. Strobel: Authorship and content analysis of engineering education research: A case study, Int. J. Eng. Pedagogy **6**(2), 39–51 (2016)
7.97 J.A. Comins, T.W. Hussey: Detecting seminal research contributions to the development and use of the global positioning system by reference publication year spectroscopy, Scientometrics **104**(2), 575–580 (2015)
7.98 Z. Liu, Y. Yin, W. Liu, M. Dunford: Visualizing the intellectual structure and evolution of innovation systems research: A bibliometric analysis, Scientometrics **103**(1), 135–158 (2015)
7.99 N. Olson, J.M. Nolin, G. Nelhans: Semantic web, ubiquitous computing, or internet of things?

A macro-analysis of scholarly publications, J. Doc. **71**(5), 884–916 (2015)

7.100 J. Pollack, D. Adler: Emergent trends and passing fads in project management research: A scientometric analysis of changes in the field, Int. J. Proj. Manag. **33**(1), 236–248 (2015)

7.101 K. Boyack, K. Börner, R. Klavans: Mapping the structure and evolution of chemistry research, Scientometrics **79**(1), 45–60 (2008)

7.102 M. Olczyk: Bibliometric approach to tracking the concept of international competitiveness, J. Bus. Econ. Manag. **17**(6), 945–959 (2016)

7.103 R.T. Li, K.A. Khor, L.G. Yu: Identifying indicators of progress in thermal spray research using bibliometrics analysis, J. Therm. Spray Technol. **25**(8), 1–8 (2016)

7.104 Y. Wang, X. Zhai, C. Liu, N. Wang, Y. Wang: Trends of triple negative breast cancer research (2007–2015), Medicine **46**, 5427 (2016)

7.105 O. Tabatabaei-Malazy, A. Ramezani, R. Atlasi, B. Larijani, M. Abdollahi: Scientometric study of academic publications on antioxidative herbal medicines in type 2 diabetes mellitus, J. Diabetes Metab. Disord. **15**(1), 1–8 (2016)

7.106 W.M. Sweileh, S.W. Al-Jabi, A.F. Sawalha, S.H. Zyoud: Bibliometric profile of the global scientific research on autism spectrum disorders, SpringerPlus **5**(1), 1480 (2016)

7.107 W.M. Sweileh, N.Y. Shraim, S.W. Al-Jabi, A.F. Sawalha, B. Rahhal, R.A. Khayyat, S.H. Zyoud: Assessing worldwide research activity on probiotics in pediatrics using Scopus database: 1994–2014, World Allergy Organ. J. **9**(1), 25 (2016)

7.108 Y. Wang, Q. Wang, R. Zhu, C. Yang, Z. Chen, Y. Bai, M. Li, X. Zhai: Trends of spinal tuberculosis research (1994–2015), Medicine **95**(38), e4923 (2016)

7.109 S. Bazm, S.M. Kalantar, M. Mirzaei: Bibliometric mapping and clustering analysis of Iranian papers on reproductive medicine in Scopus database (2010–2014), Iran. J. Reprod. Med. **14**(6), 371–382 (2016)

7.110 E.M. Gheno, D.B. Rosemberg, D.O. Souza, L. Calabró: Zebrafish in Brazilian science: Scientific production, impact, and collaboration, Zebrafish **13**(3), 1–9 (2016)

7.111 O. Tabatabaei-Malazy, R. Atlasi, B. Larijani, M. Abdollahi: Trends in publication on evidence-based antioxidative herbal medicines in management of diabetic nephropathy, J. Diabetes Metab. Disord. **15**(1), 1 (2015)

7.112 J. Li, A. Hale: Output distributions and topic maps of safety related journals, Saf. Sci. **82**, 236–244 (2016)

7.113 P. Vezyridis, S. Timmons: Evolution of primary care databases in UK: A scientometric analysis of research output, BMJ Open **6**(10), e012785 (2016)

7.114 A. Zaharia, G. Popescu, L.O. Vreja: Energy scientific production in the context of the green development models, Econ. Comput. Econ. Cybern. Stud. Res. **50**(4), 151–168 (2016)

7.115 L. Waltman, N.J. van Eck, E.C.M. Noyons: A unified approach to mapping and clustering of bibliometric networks, J. Informetr. **4**(4), 629–635 (2010)

7.116 N.J. van Eck, L. Waltman: Bibliometric mapping of the computational intelligence field, Int. J. Uncertain., Fuzziness Knowl. Based Syst. **15**(5), 625–645 (2007)

7.117 A. Rip, J.P. Courtial: Co-word maps of biotechnology: An example of cognitive scientometrics, Scientometrics **6**(6), 381–400 (1984)

7.118 M. Zitt, E. Bassecoulard, Y. Okubo: Shadows of the past in international cooperation: Collaboration profiles or the top five producers of science, Scientometrics **47**(3), 627–657 (2000)

7.119 I. Borg, P.J.F. Groenen: *Modern Multidimensional Scaling: Theory and Applications* (Springer, New York 2005)

7.120 G.S. Davidson, B.N. Wylie, K.W. Boyack: Cluster stability and the use of noise in interpretation of clustering. In: *Proc. IEEE Symp. Inf. Vis. 2001 (INFOVIS'01)*, Vol. 23 (2001)

7.121 R. Klavans, K.W. Boyack: Identifying a better measure of relatedness for mapping science, J. Am. Soc. Inf. Sci. Technol. **57**(2), 251–263 (2006)

7.122 A. Kopcsa, E. Schiebel: Science and technology mapping: A new iteration model for representing multidimensional relationships, J. Am. Soc. Inf. Sci. **49**(1), 7–17 (1998)

7.123 N.J. Van Eck, L. Waltman: *VOSviewer Manual* (Univ. Leiden, Leiden 2013) pp. 1–28

Jose A. Moral-Munoz

Department of Nursing and Physiotherapy
University of Cádiz
Cádiz, Spain
joseantonio.moral@uca.es

Jose A. Moral-Munoz received a PhD degree from the University of Granada, Spain, in 2016. He is currently working as an Assistant Professor in the Department of Nursing and Physiotherapy at the University of Cádiz, Spain. His research interests include science mapping analysis, bibliometrics, citation analysis, scientometrics, research evaluation, and data visualization.

Antonio G. López-Herrera

Department of Computer Science and Artificial Intelligence
University of Granada
Granada, Spain
lopez-herrera@decsai.ugr.es

Antonio G. López-Herrera is an Associate Professor in Computer Science and Artificial Intelligence at the University of Granada, and Vice-Dean for Academic Organization and Quality at the Faculty of Library and Communication Studies. His research interests include fuzzy information retrieval systems, fuzzy recommender systems, social media, and bibliometrics.

Enrique Herrera-Viedma

Department of Computer Science and
Artificial Intelligence
University of Granada
Granada, Spain
viedma@decsai.ugr.es

Enrique Herrera-Viedma is a Professor in Computer Science and AI at the University of Granada, and Vice-President for Research. His research interests include fuzzy decision making, bibliometrics, recommender systems, and social media. His h-index is 53 (>12500 cites WoS) and he is a Highly Cited Researcher according to Clarivate Analytics in the categories of Engineering and Computer Science from 2014.

Manuel J. Cobo

Department of Computer Science and
Engineering
University of Cádiz
Algeciras, Spain
manueljesus.cobo@uca.es

Manuel J. Cobo received his MSc and PhD degrees in Computer Sciences from the University of Granada, Spain, in 2008 and 2011, respectively. He is currently an Associate Professor with the Department of Computer Science and Engineering, University of Cádiz, Spain. His research interests include science mapping analysis, bibliometrics, scientometrics, text mining, graph mining, data visualization, quality evaluation, decision making, and recommender systems.

8. Creation and Analysis of Large-Scale Bibliometric Networks

Kevin W. Boyack, Richard Klavans

In the more than a decade since the last *Handbook of Quantitative Science and Technology Research* [8.1] was published, a sea change has occurred in the creation and analysis of bibliometric networks that describe the Science & Technology (S&T) landscape. Previously, networks were typically restricted in size to hundreds or thousands of objects (papers, journals, authors, etc.) due to lack of data access and computing capacity. However, recent years have seen the increased availability of full databases, increased computing capacity, and development of partitioning and community detection algorithms that can work effectively at large scale. As a result, much larger networks–comprised of millions or tens of millions of objects–are being created and analyzed. These large-scale networks have enabled analyses that were simply not possible in the past, analyses that require the context of complete networks to give accurate results.

In this chapter, we focus on large-scale, global bibliometric networks, and on the types of analysis that are enabled by these networks. We start by providing a historical perspective that sets the stage for recent advances that have culminated in the ability to create and analyze large-scale bibliographic networks. We then discuss data sources and the methods that are commonly used to create large-scale networks. We review many of these networks, along with the types of unique analyses that they enable, and ways that their results can be effectively communicated. After reviewing the state of the art, we discuss our most recent large-scale topic-level model of science in detail as an example of a global bibliometric model and show how it can be used for various applications.

8.1 Fundamentals and Scope

Science is a complex system that is comprised of researchers that create scientific and technical knowledge. Researchers typically work for institutions (universities, companies, etc.) and author articles that are published in journals or conference proceedings. Various metadata such as titles, abstracts, keywords, etc. associated with these articles are indexed in databases. Networks intended to elucidate the structure of science have been created from all of these features. For example, the network of articles can be used to partition science into topics and specialties. Journal networks are often used to identify fields and disciplines. Author networks are used to identify communities, and terms (i. e., keywords or words and phrases extracted from titles and abstracts) are used to create networks of concepts. Each network provides a representation of the structure of science. Often, networks created from different metadata give similar views of science, while at other times different networks reveal different or complementary features.

Scientific networks can be global or local. A local network is one that provides a representation of a particular discipline, specialty, or closely linked set of topics. In contrast, we define a global network as one that provides a representation of all of science or technology. Thus, global networks can only be based

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_8

on comprehensive data sources. Studies of local networks far outnumber studies of global networks for the simple reason that creation of a global network requires electronic access to an entire database that covers all of science, such as the Web of Science (WoS) or Scopus, or to some other very large database such as PubMed. Few researchers have this type of access. However, many researchers have institutional access to these same databases through the Internet, and can download and analyze small (local) portions of these global data. Local networks are very useful for providing some types of information, such as the dominant topic structure within a specialty, or the top researchers and institutions within a specialty. However, local networks cannot answer questions that require a larger context. For instance, a local network simply cannot identify science outside the specialty that could prove disruptive to the science within the specialty, nor can it identify which science within the specialty is–to use current buzzwords–the most interdisciplinary, translational, or transformative. Local networks cannot identify the most emergent topics across all of science. These are all boundary conditions, and without including information from outside the specialty, one cannot say where the boundaries really are. Global data are required for these types of problems.

Many studies that create and analyze local or global scientific networks use clustering or partitioning approaches. There are myriad approaches; many different similarity types and clustering algorithms are used. The full scientific network is robust at high levels of granularity [8.2, 3]–for example, physics is always closely linked to chemistry. Detailed networks are less robust. We acknowledge the fact that different methods give results that differ in detail [8.4], and that the results from any single model are a subset of many possible results. Understanding the breadth of possible results and why they occur is a very important question that has not yet been answered in the literature, and thus will not be addressed here. Despite the breadth of possible results, our experience is that the different network representations that result from different methods provide perspectives that are interpretable in their details, and that can be useful to decision makers.

In this chapter, we do not consider analyses of small-scale or local networks, but instead focus on large-scale, global bibliometric networks, and on the types of analysis that are enabled by these networks. We start by providing a historical perspective that sets the stage for recent advances that have culminated in the ability to create and analyze large-scale bibliographic networks. We review many of these networks, along with the types of unique analyses that they enable, and ways that their results can be effectively communicated. After reviewing the state of the art, we discuss our most recent large-scale topic-level model of science in detail as an example of a global bibliometric model and show how it can be used for various applications.

8.2 Background

In this section, we provide a historical perspective that suggests why one should consider using global bibliometric data to model the structure and dynamics of science. In addition, we discuss data sources, methods, and algorithms that are commonly used to create and partition global networks from these data.

8.2.1 Historical Perspective

The nature, structure and evolution of scientific networks has been a topic of interest for many decades, particularly from the perspective of citation networks. Early theory was stimulated by the desire to identify emerging topics and was centered around the notion of a 'research community'. The importance of these two concepts–identification of hot or emerging topics and delineation of research communities–to the development of methods to create and explore bibliometric networks cannot be overstated. Not only were these two of the original goals of citation analysis, but they have also formed the motivational basis for much of the work done in scientometrics over the past 60 years.

Eugene Garfield's article that introduced citation indexes [8.5] provides a clear description of the first issue–identification of emerging topics. Garfield found that subject indexes (i. e., keywords) were poor sources for identifying an emerging scientific topic as it was emerging, even if many of the key studies in that topic were published within a single journal. He found, however, that citations had the potential to track the emerging topic, and as a result he introduced the first citation index. As the citation index grew, he was able to use citation data to understand the history of key scientific breakthroughs, such as the development of DNA [8.6]. Several years later *Henry Small* introduced cocitation clustering as a way to study the specialty structure of science [8.7]. Given the computing resources available in the early 1970s, the technique was limited in practice to clustering of only the most highly cited articles [8.8]. When applied to a full database such as the Science Citation Index, cocitation clustering using only the most highly cited articles–for instance, the top 1% highly cited–produces clusters that are analogous to hot topics. This specific methodology–cocitation clustering of the most highly cited papers–has remained relatively unchanged since that time. In fact, the Research Fronts reports published annually by *Clarivate Analytics* [8.9] have used this methodology since their inception. Numerous studies using both small and large document sets have been performed specifically to identify emerging topics; see the literature review in [8.10] for a more detailed treatment. Today, identification of hot or emerging topics remains a primary concern of funders and research institutions [8.10, 11].

The notion of a research community and the desire to delineate such communities also goes back some 60 years. *Thomas Kuhn*'s [8.12] conception of a research community was of a group of about 100 researchers working on a common (empirical) problem, defined within the context of detecting scientific revolutions (or identifying emerging topics!) as they were occurring. *Nicholas Mullins*' [8.13] study of the research communities in sociology revealed a few research communities of approximately this size, along with a host of much smaller communities. Subsequent work by *Diana Crane* [8.14] also suggested that there would be roughly 100 people per research community. Both Mullins and Crane viewed the evolutionary paths of these communities in a similar fashion as follows. One starts with a large number of very small research communities that are weakly organized. A major discovery in a community–a relatively infrequent event–results in a thickening of the network relationships in that community. With additional discoveries, the research community attracts new members and grows larger until it is comprised of around 100 researchers. As the community ages, the rate of discovery drops and the community size stabilizes. As the rate of scientific or technical advance decreases further, the size of the community declines as researchers migrate to other communities and research problems.

A model of the scientific literature that is based on global bibliometric data fits very well with this conceptualization of research communities and also with the notion of topics. If one defines a topic as a collection of documents with a common intellectual interest, or as a particular problem or closely related set of problems in science, one can also define a research community as the group of researchers working in that topic, and the two ideas can be roughly equated. Topics/communities can be new or old, large or small, and growing or declining in terms of their membership. Each topic has a story, and it is that story that is of interest to the various stakeholders in the science system, whether they be funders, administrators, or researchers. We seek methods that allow us to accurately identify topics and their dynamics that will enable us to tell the stories associated with topics. Global data is critical to this endeavor. While some topics identified using small (i. e., local) datasets, such as those created from journal sets or keyword searches, are complete and relatively accurate, other topics based on local data can be incomplete or even misleading in terms of the stories they tell [8.15, 16].

Of course, our understanding of topics can also be incomplete if we do not understand the larger context within which scientific and technical work takes place. Scientometrics has historically focused its attention on the analysis of scientific papers, one of the key outputs of the science system. Although this is understandable from the standpoint that literature data is relatively available, it is time that the scientometrics community starts to investigate, incorporate, and link other data sources into their thinking and analyses.

As an example of the types of data that could be considered for bibliometric analysis, and more particularly for large-scale analysis of bibliometric networks, we refer to Latour's model of the research laboratory in his book *Science in Action* [8.17]. This is a much more complex view of the science system than is indicated by publications alone (Fig. 8.1). Latour proposed a positive feedback loop in which one starts with new discoveries that are published (*argued*) in the scientific and technical literature. These discoveries give rise to *innovations*, which are rewarded by more funding (*money*) that can be spent on salary (*workforce*) and *instrumentation* that are used to do more research and create laboratory *objects* or artifacts. The cycle then continues, with researchers presenting their newer findings in the scientific and technical literature, etc. The model represents a positive feedback loop where growth occurs as long as current findings are considered to be significant contributions. Although Latour's model was based on observations of how research laboratories grow, we view it more generally–it is intuitive to think of it as a conceptual model of the evolution of topics or research communities. We note that Latour's model only shows movement from smaller to larger circles. In contrast, we note that the activity in some topics can and does decline over time, and thus expand upon the model in that way.

The balance of this chapter will build on the perspectives developed above. We suggest that large-scale bibliographic networks are needed to accurately identify topics in the scientific (and technical) literature. What is large scale? For the purposes of this chapter, we define a large-scale data source as one that is relatively comprehensive in terms of coverage and thus provides

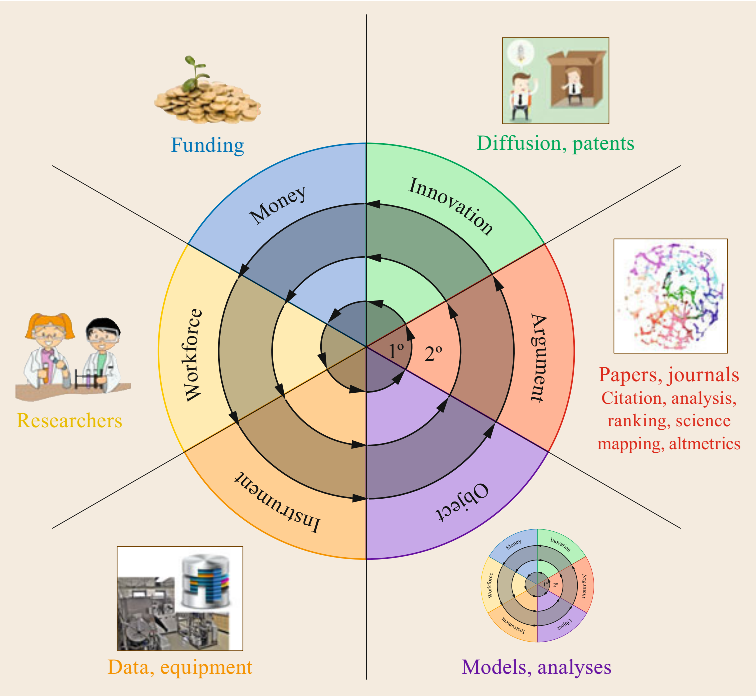


Fig. 8.1 Latour's model of the process of science, overlaid with examples of features we associate with each portion of the process. Many of these features, alone or in combination, are amenable to network analysis (after [8.17, p. 160])

Table 8.1 Large-scale bibliometric data sources

Source type	Source	# Records	Availability
Scientific literature	Scopus (1980–2016)	50 M	Subscription
	WoS (1980–2016)	47 M	Subscription
	PubMed (1980–2016)	22 M	Open
	Microsoft Academic Graph	130 M	Open
	CrossRef	90 M	Open
	Chemical Abstracts	24 M	Subscription
	JSTOR	9 M	Subscription
	CAB Abstracts	8.6 M	Subscription
Full-text literature	ScienceDirect	9 M	Subscription
	IEEE Xplore	4 M	Subscription
	CiteseerX	2 M	Open
	PubMed Central Open Access	1.5 M	Open
	arXiv	1.2 M	Open
Patents	Espacenet	> 90 M	Subscription
	PatStat	> 90 M	Subscription
	Derwent	> 90 M	Subscription
	LexisNexis	> 90 M	Subscription
	USPTO (1980–2016)	5.9 M	Open
	PatentsView	5.9 M	Open
Project-level funding	StarMetrics	0.8 M	Open
	UberResearch	> 1 M	Subscription

a reasonable representation of either all of science or technology or a deep representation of a high-level field (such as physics, computer science, or medicine). Further, to be a large-scale network, we suggest that the network created from the data must be granular and detailed. The bibliometric networks that we will consider here will, for the most part, be comprised of at least 1 000 000 documents and will have thousands or tens of

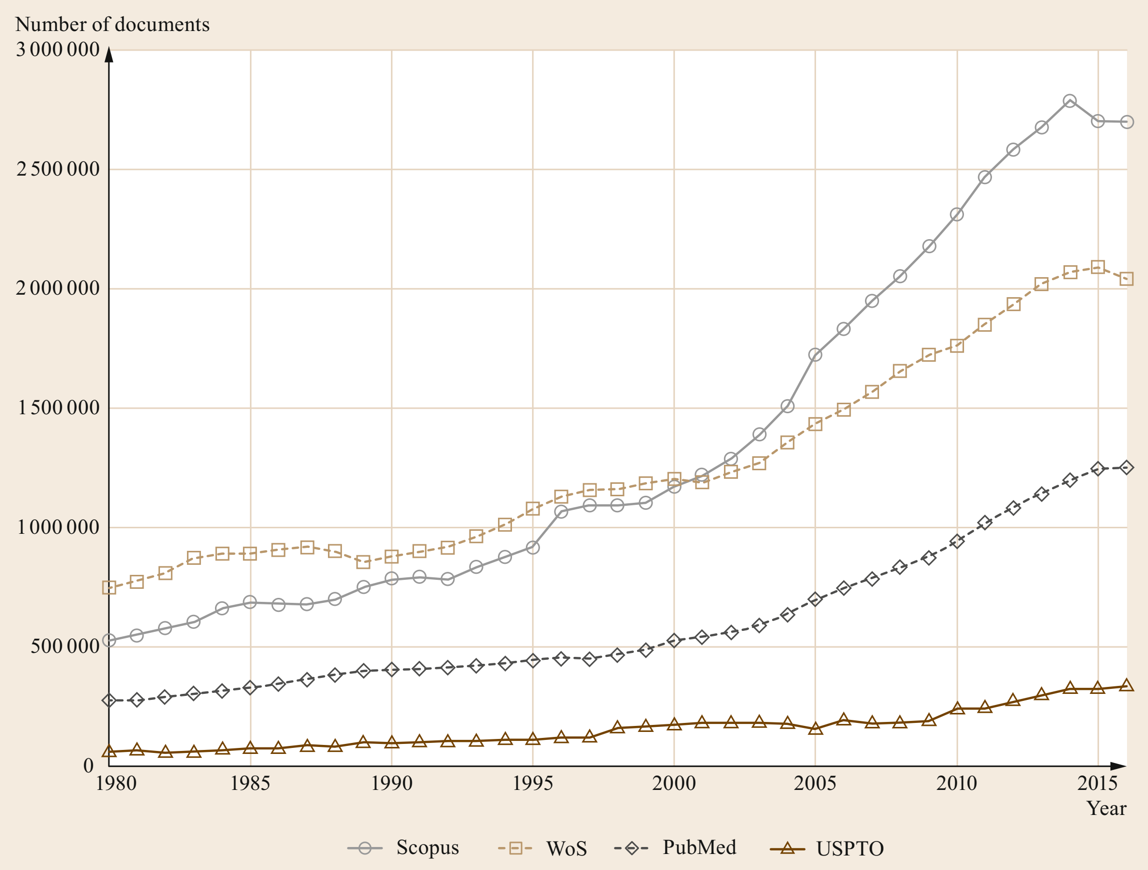


Fig. 8.2 Numbers of records by year for four prominent databases

thousands of clusters. We further suggest that we should expand the types of data (and thus networks) that we analyze to include funding, researchers, equipment, and laboratory artifacts. We realize that these data have historically been less available than publication and patent data, but this does not mean that we should simply dismiss the possibility that these data will help us better elucidate the stories associated with the science system to the benefit of all stakeholders. Rather, we suggest that we should attempt to procure more of these data for analysis so that we can better understand the science system.

8.2.2 Data Sources

There are many data sources that can be used to create large-scale bibliometric networks. Table 8.1 lists a number of these, along with approximate numbers of records and the availability of each source. This list is certainly not complete, but does cover the majority of large-scale sources that have been used for bibliometric studies. Each source will be described briefly here.

Scientific Literature

The two primary scientific literature databases used for most scientometric and bibliometric studies are Scopus (Elsevier) and the Web of Science (WoS, Clarivate Analytics). These are the dominant databases because they are the only two that cover (more or less) all scientific fields and include cited references, thus enabling citation analysis and the construction of citation-based networks. WoS is descended from Garfield's original Science Citation Indexes and was the only citation database available until 2004 when Scopus was introduced. Among the two databases, Scopus has broader recent coverage (since 2001; Fig. 8.2) while WoS has far more historical coverage. Both databases require a subscription, and access to either full database in electronic form typically requires a specific license.

PubMed and Medline are literature databases that are maintained by the US National Library of Medicine (NLM). Medline primarily indexes publications from the life and medical sciences, while PubMed includes all of Medline along with select publications (although not entire journals) from outside those fields. PubMed does not include cited references, but does include MeSH (Medical Subject Headings) terms, chemicals, accession numbers, and other data types that link to other databases maintained by NLM. Data from these fields, along with the standard textual fields (title and abstract), can be used to create co-occurrence networks between PubMed records. The majority of PubMed records are also available in Scopus and WoS, with Scopus claiming 100% PubMed coverage. PubMed has the advantage of being an open database that can be downloaded freely for scientometric analysis.

Microsoft Academic Graph is another large-scale resource that has recently become available. As suggested by the word *graph* in its name, it includes cited references that enable creation of a large-scale bibliometric network. Although it contains around 130 million records, only 35–50 million of those records are linked by citation [8.18, 19]. Nevertheless, it is free for research purposes (but not for commercial use), which makes it a very valuable resource for the creation and exploration of large-scale networks. CrossRef, in addition to being a Digital Object Identifier (DOI) registration agency, maintains metadata associated with DOIs that can be downloaded for analysis. We do not list Google Scholar in Table 8.1 because, although it is perhaps the largest bibliometric resource in the world (estimated at 160 million records), it has not been made available in bulk for scientometric analysis.

The other three scientific literature databases listed in Table 8.1 each focus primarily on a single large field–Chemical Abstracts on chemistry, JSTOR on the social sciences, and CAB Abstracts on agriculture. Each is subscription-based, and none has been used extensively for scientometric analysis, but could conceivably be used to create and analyze a large-scale network of topics in those fields. Many other literature databases exist, but they are typically much smaller than those listed and are either subject-specific or region-specific, and will not be covered here.

Full-Text Literature

In addition to those databases containing literature metadata, several vendors and services now maintain databases with large numbers of full-text documents. None of these yet has the type of broad and deep coverage offered by either Scopus or WoS. Nevertheless, we can expect access to full-text documents to increase in the future, and as such, they have the potential to become significant sources for large-scale bibliometric networks and analysis. Table 8.1 lists ScienceDirect (Elsevier) and IEEE Xplore, each of which houses millions of full-text documents. ScienceDirect is broad, covering most fields of science, while IEEE covers mainly electrical engineering and computer science. Although not listed in Table 8.1, Springer, Wiley, and other publishers are also making full text available to their subscribers, and many of these publishers (including Elsevier, Springer, and Wiley) allow downloading of their documents in XML format. Full-text data from CiteseerX, arXiv, and PubMed Central are all freely available through their APIs (application programming interfaces), and can be mined for references, citation contexts, and other features that can be used to create networks.

Patents

Patents are another type of output from science and technology. We list patent sources in Table 8.1 for completeness, but will not elaborate much on patent networks in this chapter since they will be covered in detail in other chapters. Nevertheless, there are multiple vendors (Espacenet, PatStat, Derwent, and LexisNexis) that collect and unify patent data from jurisdictions around the world. It is difficult to determine large differences in these sources without detailed examination of each, so we have listed them as equivalent in terms of coverage in Table 8.1. Although there are undoubtedly differences between vendors, the bulk numbers of worldwide patents covered by each is likely to be roughly similar. Patent data contain references to prior art, including prior patents, patent applications, and scientific papers. Links to scientific papers, known as nonpatent references (NPR) are typically very dirty data, and require cleaning and linking to literature databases to be used properly. We are unaware of any public source of cleaned patent-to-paper linkage data, but are aware of institutions that have done this work and maintain these links for their own use [8.20, 21].

Patents are also assigned to categories in one or more classification systems. Both feature types (links and categories) can be used to link patents into large networks. Among free sources, we list data from the US Patent & Trademark Office (USPTO) and PatentsView, a source for cleaned and unified USPTO data that has been funded by the US National Science Foundation (NSF) for many years to process these data and make them more widely available.

Project-Level Funding

Data on the funding of science are perhaps the hardest data to obtain for many reasons. Although aggregate data are available in reports from, for example, the Organisation for Economic Co-operation and Development (OECD) and US National Science Board, project-level data are needed to really understand the effects of funding on science and technology at the topic level. Private funding by companies or foundations is, of course, proprietary to the funding institutions. Few public institutions, such as government agencies, have made their project-level data publicly available. Perhaps the largest repository of funding data currently available is UberResearch, a relatively young company that gathers project-level funding data from sources around the world and makes analysis based on those data available to subscribers.

Among open data sources, the largest dataset, in terms of numbers of projects and funding amounts, has been made available for bulk download by the US StarMetrics consortium. These data contain project-level metadata including annual funding amounts since 2008 for many US agencies including the National Institutes of Health (NIH), NSF, and the National Aeronautics and Space Administration (NASA). Abstracts are available for most projects, and the data also include links from PubMed articles to NIH projects mined from PubMed publications and internal NIH source materials. Links are not yet available for projects from other agencies, but will hopefully be added in the future. Project-level data (including links) for NIH for years prior to 2008 are available through the NIH RePORTER and ExPORTER websites, while earlier data for NSF are available through an API.

Data that link specific projects and scientific articles are also available from WoS, Scopus, and PubMed. WoS began indexing acknowledgments in 2008, and has indexed nearly 11.7 million specific grant-to-article links through 2016. Scopus also indexes acknowledgments; however, they started later, and the number of annual links indexed by Scopus is only about one third of those indexed by WoS. PubMed also lists grant numbers; however, they do not keep all information, but limit their listings to NIH grants and select grants from a few other funders such as CDC and the Wellcome Trust. In general, the grant-to-article data from PubMed are largely redundant with those available from NIH ExPORTER.

We note that there are a couple of other smaller sources of project-level funding data that are publicly available. Cordis makes available files with data from 40 000 grants from the FP6, FP7, and Horizon 2020 EU funding cycles. The UK Gateway to Research (GtR) lists perhaps 70 000 grants with metadata and articles reported to have acknowledged those grants. We have linked the GtR data to Scopus articles, and have posted the link file for download at http://www.mapofscience.com/wp-content/uploads/2016/05/links_pub_gtr.xlsx.

All of the data sources listed above are suitable for creating large-scale bibliometric networks. Perhaps more interestingly, we look forward to the day when data of multiple types–papers, patents, funding, and perhaps other things like equipment–will be linked together to form more detailed and accurate pictures of science, technology, and innovation.

8.2.3 Methods

The process that is used to create a bibliometric network, whether small or large, is relatively simple conceptually, and consists of the following general steps:

1. Data extraction
2. Similarity calculation
3. Clustering
4. Layout and visualization.

In some cases, there can be multiple iterations of the similarity calculation and clustering steps. Despite this difference in detail, the four steps listed here are the major building blocks of any network creation and/or visualization. Each will be discussed further here. An alternate, and more general, explanation of these steps that is applicable to both small and large networks can be found in [8.22].

Data Extraction

Bibliometric networks, like all other networks, consist of objects and the links between those objects. When speaking of a network, we often refer to these as nodes (objects) and edges (links). The first steps in creating a bibliometric network are to decide which data to use, what the nodes will be, and what feature or features will be used to create the edges. For networks based on large-scale bibliographic data, either a full database or a substantial fraction of that database is typically used. For instance, one can use the full Scopus database (tens of millions of articles) as the data source, or one can use a one-year slice of the database (2–3 million articles) as the basis for the network.

The choice of what to use as the network nodes depends on the question that is being asked and the type of network that can be used to answer that question. If the question is about how authors link together, then a collaboration network based on coauthorship is good choice. If the question is about the relationship between concepts, then a network based on the relationship between terms or keywords is a good choice. If the question is about topics and their dynamics, then a network of topics, which we have previously defined as collections (or clusters) of documents with common intellectual interest, is a good choice. The balance of this chapter addresses questions that are best answered from a topic perspective, and we will thus move forward with this example.

For a topic network, we ultimately define nodes as topics or clusters of documents, and edges as either citation or textual links between topics. However, since topics are not available in any database, we must create topics from documents. Topics result from clustering a large-scale bibliometric network of documents. For our example network, we thus define articles as nodes, and we will define the edges as citation-based relationships between the nodes. Note that we could define the edges based on any type of valid relationship between documents. Examples include co-occurrence of textual features in the titles and abstracts of documents or co-occurrence of authors across documents. We choose to use citation relationships because they are typically intentionally chosen by authors, and in our experience are less interchangeable than words.

Having chosen to use documents as nodes and citations to create the edges of our network, we make sure that our data extractions include these pieces of information. In this case, a list of the edges (pairs of citing and cited document IDs) is our starting point, because the list of edges necessarily contains all of the nodes.

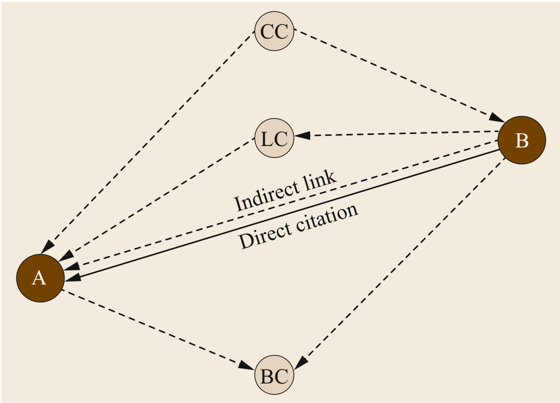


Fig. 8.3 Citation links can be used to create several different types of edges that can be used to create bibliometric networks (after [8.23]). Indirect link types include CC (cocitation), BC (bibliographic coupling), and LC (longitudinal coupling)

Similarity Calculation

The next step in creating a bibliometric network is to calculate the similarity between pairs of nodes in the network. This is not as trivial as it may seem. Of course, for our example network comprised of documents and the citation relationships between documents, one could take the simplest approach and simply designate each citing-cited document pair as an edge, and weight each edge equally. This is the simplest approach, known as direct citation, and it can work quite well. However, citation data enable more sophisticated similarity measures as well. While direct citation is a first-order link between two documents, indirect links can also be created and used as the basis for similarity between documents. Figure 8.3 shows three such linkage types–cocitation (CC), which occurs when a document cites documents A and B; bibliographic coupling (BC), which occurs when documents A and B both cite the same document; and longitudinal coupling (LC), which occurs when A cites a document that cites B. Each of these similarity types can be generated from the same set of initial edges, forming a new derivative set of edges that is typically much larger than the original set of edges. The different edge types can also be combined as was demonstrated by *Small* [8.23].

Figure 8.3 exemplifies edges that are based on simple counts of links. While direct citation links between pairs of documents are singular (except in rare cases), indirect links between the same pair of documents can occur many times. For example, a pair of documents that are cocited together in many subsequent papers will have a large number of such indirect links, and can thus be said to have a large cocitation score. These raw scores can be used to weight the edges in a large network. However, experience and experimentation have shown that in most cases, normalizing the raw count scores leads to a more accurate clustering of the network. Thus, edges with normalized weights are typically used to create bibliometric networks.

Many common normalization methods are available; different normalizations are used for different types of similarities. For edges based on direct relationships, such as direct citation edges, normalization is typically based on the number of edges for each node. One method assigns each edge a weight of $1/n$ where n is the number of edges leaving the node [8.24], for example, for scientific documents n would be the number of cited references in the document. Another method assigns edge weights based on edges entering and leaving the node [8.25].

For edges based on co-occurrence relationships, including cocitation and bibliographic coupling, coterm analysis and coauthor analysis, most normalization approaches are based on the numbers of edges related to both nodes. The cosine, Jaccard, and inclusion indexes, and the association strength, are all commonly used and their relative strengths and weaknesses were recently compared [8.26]. Historically, the choice of which of these measures to use seems to have been personal rather than driven by any particular evidence. Regardless of which choice is made, for this step a list of edges with weights is calculated from the list of edges generated from the data extraction step, and this list of weighted edges is then used as input to the clustering step. At this point, for completeness, we note that although we typically use citations to estimate the relatedness between document pairs, others prefer to use textual characteristics. There is relatively little difference between the accuracies of the best text-based and citation-based relatedness measures when it comes to clustering documents [8.24, 27]. However, computation requirements for citation-based relatedness measures are typically far less than for text-based measures.

To continue our example, in this step we choose to use direct citation as the basis for our edges, and will normalize the edges by the number of references per citing document. Our edge file thus contains all pairs of citing and cited document IDs along with edge weights that are present in our dataset.

Clustering

Now that we have an edge file that represents a very large network of documents, it is time to partition the overall network into groups of documents that represent topics using a clustering algorithm. We do not undertake a history of clustering algorithms here, but mention only those that are currently capable of clustering tens of millions of objects and that are somewhat familiar to bibliometric researchers. A detailed review of clustering or community detection techniques can be found in [8.28].

There are several classes of algorithms that can cluster a very large network (at least one million nodes and ten million edges), including modularity, label propagation, spectral analysis, and map equation algorithms. It is difficult to compare the results from different clustering algorithms because a valid comparison is best made when the cluster size distributions are very similar. A recent study comparing clustering results using algorithms from each of the classes listed here on data sets of 1.2 million and 11 million nodes was inconclusive because the cluster size distributions from the different classes were quite different [8.29]. However, the authors did conclude that map equation algorithms performed best when comparing results on smaller datasets focused on bibliometrics and information science. They thus advocate that map equation algorithms should receive more attention than they currently do. These recommendations could not be extended to larger networks.

Each class of algorithm has its advocates. For instance, map equation algorithms, such as InfoMap [8.30], are used on very large networks by Jevin West and associates at the University of Washington [8.31], while modularity-based algorithms include the well-known Louvain algorithm [8.32] and the smart local moving (SLM) algorithm [8.33] developed at Leiden University. We personally favor the SLM algorithm, and its VOS predecessor [8.25], because they are highly versatile and can be tuned to generate networks with different numbers of clusters. When partitioning large networks into large numbers of clusters, our experience is that the cluster size distributions are relatively flat, spanning only two orders of magnitude in size.

To summarize this step in our process, although details differ, most clustering algorithms accept lists of weighted edges in some form as input, and return a list of the cluster to which each node is assigned as an output. To continue our example, using a list of direct citation weighted edges as input, and using the SLM algorithm for clustering, our output would be the cluster assignments for each document. If we set the input parameters such that the documents were partitioned into roughly 100 000 different clusters, each cluster would represent a single topic in science.

Layout and Visualization

The final step in our general process is to create a visual map of the resulting network. In some cases, this step is not needed. If the research question does not require visualization, the cluster solution resulting from the first three steps is sufficient, particularly if the cluster contents and their metadata are maintained in a database that can be queried to answer questions. However, we find that visualization is helpful in many regards to disseminate results, and is thus applicable to most network analyses.

Visual maps of networks can be created with many different tools. Several of these tools combine the similarity calculation, clustering, and visualization steps, leaving only the data extraction to the user. For example, CiteSpace [8.34, 35] inputs a set of records downloaded from WoS or other data sources, and creates a number of different visualizations based on cocitation clustering. VOSviewer [8.36] is widely used to create useful visualizations of documents, journals, terms, or authors from records downloaded from either WoS or Scopus, and clusters the data as well as providing a visual rendering. Several other such tools exist [8.37, 38]. However, they are all designed to work with document-level records–with data directly obtained from WoS, Scopus, or another database–and are not well equipped to work with cluster-level data that has already been created using the process mentioned above. These tools are also limited in the size of the dataset they can process and visualize.

Layout (i. e., assigning coordinates to each node) and visualization of a network of clusters are best accomplished using other tools. Network layout algorithms require nodes and edges as input. Thus, to visualize cluster-level data, an additional step must be taken here–similarity between pairs of clusters must be calculated. Cluster-to-cluster similarity can be calculated in a variety of ways as mentioned above in Sect. 8.2.3, *Similarity Calculation*. It can be based on aggregated citation relationships between clusters, or it can be based on another feature such as the text associated with the documents in each cluster. Once the cluster-level similarity has been calculated, an edge list is created where the edges consist of cluster IDs and the edge weights (or cluster-cluster similarity values). At this point one can either cluster the data again using the *clustering* step mentioned above to create higher-level clusters, or one can create a layout of the cluster-level network using a graph drawing or network layout algorithm such as Kamada-Kawai [8.39], Fruchterman-Reingold [8.40], or DrL/OpenOrd [8.41]. Fortunately, this step is not as difficult as it might sound because these algorithms are implemented in other user-friendly network analysis tools such as Pajek [8.42] and Gephi [8.43]. Statistical packages, such as R, are also capable of sophisticated network analysis and graph drawing.

Tools like Pajek and Gephi can be used to do the layout and visualization of a cluster-level network, or they can be used to simply visualize a network whose layout is calculated elsewhere, such as with a standalone version of DrL, which can handle networks of several million nodes. This is the process that we use. To continue our example, once we have created topics, or clusters of documents using direct citation edges, we assemble the text for each cluster and calculate a cluster-to-cluster similarity using the BM25 text relatedness measure [8.44, 45]. We then use DrL to create a layout of the clusters, and use Pajek to visualize the resulting network [8.46]. Figure 8.4 shows an example of the visual map resulting from the example that we have used to illustrate the process steps given here. Each topic, represented by a dot, is a cluster of papers that are linked by citations. Topics that use very similar text are near to each other in the map. Each topic in the map has been designated as belonging primarily to one of twelve high-level fields, and is correspondingly colored in Fig. 8.4. These designations were made by assigning papers to fields using the University of California San Diego (UCSD) journal-to-field assignments [8.47].

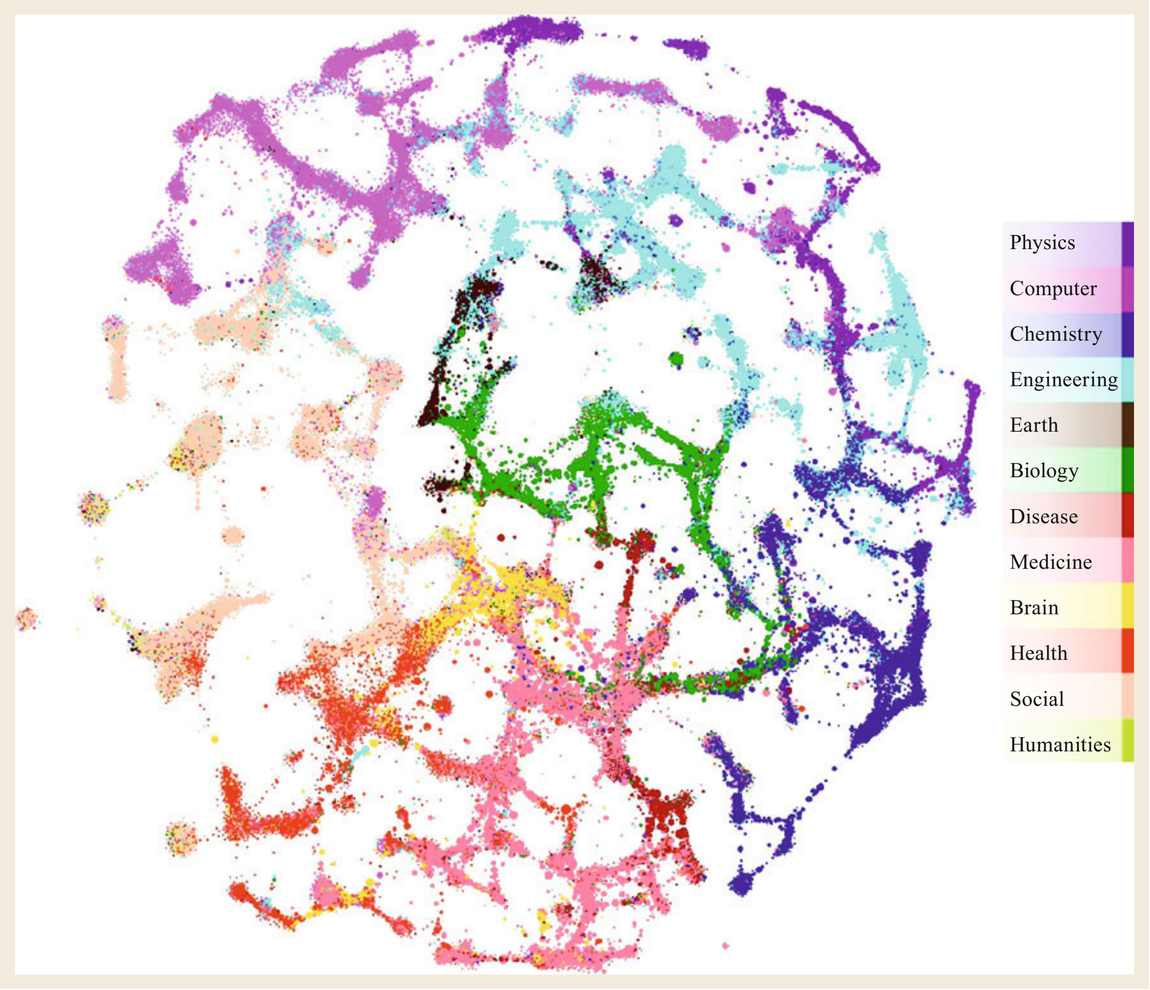


Fig. 8.4 Visual map of topics in science. Each *dot* represents a topic or research community

Part A | 8.2

8.3 Studies of Large-Scale Bibliometric Networks

Now that the general process for creating a large-scale bibliometric network has been explained, we turn to a review of recent studies of these types of networks. In addition to the work to cluster large-scale bibliometric networks, we will also review other studies that create large-scale networks but that do not necessarily cluster them.

8.3.1 Networks of Scientific Topics

The general methodology mentioned in the previous section can be used to cluster and visualize networks of all different sizes. It is not specific to large-scale networks, and in fact has been used much more often on small- and medium-sized networks (those with fewer than 100 000 documents, for instance) than on large-scale networks. There are far fewer studies that present and analyze large-scale networks, largely because few institutions have access to complete bibliographic databases in electronic form. Nevertheless, since this chapter is specifically about large-scale networks, we ignore the large body of work on smaller networks, and focus here on work based on sets of roughly one million documents or more. We also ignore journal networks here because, although some journal network studies have been based on millions of documents, the networks have ultimately been composed of either thousands of journals or hundreds of journal categories, and thus do not meet our criteria for being a large-scale bibliometric network. This section will also include studies of large-scale data seeking to establish the relative accuracy of different similarity methods because the results of such studies naturally reflect on the accuracy of the application space associated with large-scale bibliometric networks.

Clustering of documents in bibliometric studies was limited to document sets of far less than one million documents from the early 1970s until the late 2000s. Here we review those studies of one million documents or more in roughly chronological order, as shown in Table 8.2. *Boyack* [8.48] was the first to publish a study (in 2009) where a million documents were clustered into topic-level categories. Using a set of one million documents from the 2003 WoS file year and seven million bibliographic coupling links between documents, he used VxOrd (the predecessor to DrL) to create a document layout. Although the layout did not assign documents to clusters, it did concentrate them in cluster-like structures. Single-link clustering using edges and distances was then used to create clusters from the layout results, with the one million documents being assigned to 117 435 clusters, each of which represented a scientific topic. This classification was then used to identify specific potential collaborations because one can reasonably expect that authors in the same topic that have not previously coauthored together could become collaborators in the future. Using specific sets of topics in physics, these potential collaborations were aggregated at the institution level to identify those institutions that had a high potential overlap, and thus a high potential for fruitful collaboration, such as Sandia National Laboratories. This is an illustration of but one practical application of clustering documents into a large number of topics. The results would have been much less specific, and potentially much less actionable, if the documents had been clustered into fewer clusters. Specificity increases with granularity, and this specificity is crucial to many applications, such as the identification of potential collaboration.

A related process was used shortly thereafter by *Klavans* and *Boyack*, but using Scopus data, cocitation links rather than bibliographic coupling links, and the DrL layout algorithm to cluster 2.08 million reference papers that were cited in 2007 [8.49]. The cluster formation step was somewhat different in this case. DrL employs edge cutting, and is a stochastic algorithm in that changing the starting seed value changes the order in which edges are considered, and thus changes the edges that will be cut. Results from two DrL calculations on the same input file with different seeds will have highly overlapping, but not identical, sets of uncut edges. In this study, DrL was run with ten different seeds, and an edge was considered robust only if it remained uncut in six of the ten runs. Clustering was done using the uncut edges using a standard average-link clustering algorithm. Once the 84 000 clusters were obtained, 5.68 million papers from a five-year (2003–2007) subset of Scopus were fractionally assigned to the clusters using citation links. These 84 000 clusters were then combined in an idiosyncratic way by institution (i. e., universities, countries), and the results were analyzed to identify the topic-level strengths of those institutions and countries. Portfolio analysis is thus another application that benefits from a highly detailed clustering of the scientific literature.

It was at about this time that SciTech Strategies (STS) received a Small Business Innovation Research (SBIR) grant from NIH to compare the accuracies of citation-based and text-based similarity measures. A team consisting of researchers from several institutions was assembled to do this work. A dataset of 2.15 million documents that were available in both Scopus and PubMed was created, and 13 different similarity measures–three citation-based, nine text-based, and one

Table 8.2 Analyses of large-scale bibliometric networks and associated characteristics. Numbers of nodes and edges apply to the single largest clustering calculations performed in each study, and do not necessarily represent final totals

Study	Data	Unit	Aim	Sim	Alg	#Nodes	#Edges	#Clust	Uses
Boyack [8.48]	WoS	Papers	Topics	BC	VxOrd	1.00 M	7 M	117 k	Recommend collaboration
Klavans and *Boyack* [8.49]	Scopus	Papers	Topics	CC	DrL	2.08 M	20 M	84 k	Portfolio analysis
Boyack and *Klavans* [8.24]	Scopus + PubMed	Papers	Topics	DC, CC, BC, Hybrid	DrL	2.15 M	15 M	30 k +/−	Accuracy
Boyack et al. [8.27]	Scopus + PubMed	Papers	Topics	Text	DrL	2.15 M	20 M	30 k +/−	Accuracy
Klavans and *Boyack* [8.16]	Scopus	Papers	Topics	CC	DrL	2.5 M	20 M	340 k	Field and topic characterization
Boyack and *Klavans* [8.46]	Scopus	Papers	Topics	CC+BC	DrL	3.27 M	30 M	151 k	Field characterization
Waltman and *van Eck* [8.25]	WoS	Papers	Specialties	DC	VOS	10.2 M	98 M	22 k	Field characterization
Waltman and *van Eck* [8.33]	WoS	Papers	Specialties	DC	SLM	10.6 M	105 M		Algorithm development
Ruiz-Castillo and *Waltman* [8.50]	WoS	Papers	Specialties	DC	SLM	9.4 M		73 k	Research evaluation
Small et al. [8.10]	Scopus	Papers	Topics	DC	VOS	17.0 M	100 M	84 k	Emerging topics
Boyack and *Klavans* [8.51]	Scopus	Papers	Topics	EDC	VOS	43.4 M	511 M	156 k	Structure, coverage
Klavans and *Boyack* [8.52]	Scopus	Papers	Topics	EDC	SLM	48.4 M	582 M	92 k	Accuracy
Foster et al. [8.53]	Medline	Chemicals	Novel links	Co-occur	MapEq	181 k	10.5 M		Novelty
Shi et al. [8.54]	Medline	Authors, MeSH	Novel links	Co-occur	MapEq	9.3 M auth, 16k MeSh			Novelty
Wesley-Smith et al. [8.55]	MS Acad	Docs	Papers	DC	MapEq	38 M	600 M		Recommend papers
West et al. [8.56]	MS Acad	Docs	Papers	DC	MapEq	27.3 M	262 M		Recommend papers
This study	USPTO	Patents	Topics	DC	VOS	3.3 M	29 M	27 k	Portfolio analysis

hybrid (citation and text)–were calculated. The same clustering method, using the commonalities in ten DrL runs, was used for all 13 similarity measures. Clustering results were compared using the concentration of grant-to-article linkages [8.24, 27]. The working assumption behind this metric was that articles acknowledging a single grant would be topically similar, should thus be concentrated in the cluster solution. Cluster accuracy should therefore correlate positively with grant concentration. These studies showed that of the three citation-based similarities, bibliographic coupling was the most accurate, followed closely by cocitation (CC). Among text-based similarities, the PubMed related articles similarity gave the best results, followed closely by the BM25 measure. The best citation-based and text-based approaches gave roughly similar results, and the hybrid measure, which included a textual component with bibliographic coupling, did slightly better than either. Although this study did not address a specific topic-level application such as portfolio analysis, it did establish that text-based and citation-based topics could be created using millions of documents. It also showed that, while bibliographic coupling gave better results than cocitation analysis, the gap between the two was not large, and that there was not a compelling reason to give up cocitation analysis (which STS had been using for years) as the basis for creating models of science.

STS continued to use its cocitation analysis process to create detailed models of science and expanded from using single-year models, each based on more than two million reference papers, to linking together models from adjacent years into thread-like structures [8.16]. Improvements in the early 2010s included the creation of a hybrid CC-BC approach that increased cluster accuracy [8.46]. The intent behind linking annual sets of clusters was to produce a dynamic, rather than a static, view of science that would enable identification of emerging topics. A cluster from one year was linked to a cluster in the following year if the fraction of papers in common was above a certain threshold. These cosine thresholds were lower than one might expect (around 0.25), and reflect the inherent instability of cocitation sets from one year to the next. The most advanced model linked together cluster sets from 16 years, and the resulting threads or topics–those that lasted two years or more–were used to identify discipline-like structures based on documents rather than journals. These topics based on linked cocitation clusters were also used for a couple of practical topic-level applications. The first of these was to distinguish topics along the basic-to-applied spectrum using research levels [8.57]. The second was to characterize the detailed structure and dynamics of sets of topics related to emerging areas in science using research on graphene and dye-sensitized solar cells (DSSC) as two examples [8.58]. This topic-level analysis suggested that research on graphene and DSSC could not be characterized as single topics, but that each area consisted of multiple detailed topics with different dynamics, some of which were focused on fundamental (material) properties and others that were clearly application oriented. Interviews with program officers using cluster contents (on different areas than graphene and DSSC) showed that experts could easily distinguish between these detailed topics, and in many cases knew the specific stories that accompanied each topic.

Each of the studies reviewed so far, while large in scale, were also limited in that DrL has a practical limit of about three million nodes and 30 million edges on a typical computer, and thus cannot be used to cluster the entire contents of a database such as WoS or Scopus in a single operation. This barrier was overcome in 2012 and 2013 by the introduction of new modularity-based clustering algorithms by Ludo Waltman and Nees Jan van Eck at Leiden University that are capable of clustering tens of millions of documents. The first algorithm was an extension of the clustering code used in VOSviewer (VOS in Table 8.2), was written in the C language, and was first used to cluster more than ten million WoS documents connected by 98 million direct citation edges [8.25]. The VOS code is capable of clustering at multiple resolutions, and doing so in a hierarchical manner. In this first demonstration, clustering was done at three different resolutions such that solutions with 20, 672, and 22 412 clusters were created. Characterizations were done at each level of clustering, including a ranking of the hottest topics based on average publication year at the second level. This study established that a very large, multiyear set of scientific publications could be clustered in a single operation, and that at the most granular level the resulting clusters had varying dynamics and a detailed topical focus.

A second modularity-based algorithm, written in Java, and now known as the smart local moving (SLM) algorithm, was also introduced by *Waltman* and *van Eck* [8.33]. SLM was compared to the well-known Louvain algorithm and to a multilevel refinement of the Louvain algorithm. While the results of all three algorithms were very similar for a single iteration, SLM results improve with each additional iteration. Although running multiple iterations does increase the time needed to cluster, the increased modularity (which should correspond with increased accuracy) can be worth the expense. The SLM algorithm was also run on a set of nearly 40 million web pages with 783 million links, thus establishing that it is capable of clustering the complete contents of either WoS or Scopus from a size perspective.

SLM was later used for a practical application by *Ruiz-Castillo* and *Waltman*, who clustered a ten-year set of WoS documents at 12 different levels of granularity, ranging from 390 to 73 205 clusters [8.50]. The purpose of this study was to determine an optimum number of clusters to use as the basis for normalization of indicators used in research evaluation. It was determined that a solution with less than a thousand clusters merged fields that should be distinct from each other, while a solution with tens of thousands of clusters contained many clusters that were too small to be used for research evaluation. Solutions with a few thousand clusters were deemed most acceptable for the purpose of research evaluation. Accordingly, SLM has been used to create the roughly 4000 document clusters used in the Leiden Ranking for the past few years.

In 2012, we (Klavans and Boyack) were very aware of the size limitations of DrL, and had begun looking for ways to cluster larger datasets. We became aware of the VOS algorithm and results in a prepublication version of [8.25] and were excited about the possibilities. We obtained a copy of the code (which was made openly available), and quickly set about replicating the study and performing calculations to compare to our cocitation models. We found that the accuracy of the direct citation VOS cluster solution was comparable to that of our previous cocitation and bibliographic coupling calculations, which suggested that the direct citation results in our previous study on accuracy [8.24] had been adversely affected by the short time window associated with that dataset. Upon detailed examination of numbers of clusters, we became convinced that these direct citation clusters were very good representations of historical views of topics, but that they were also quite different in character from the topics produced by linked annual cocitation models. Where topics based on cocitation were only partially stable from year to year, topics based on direct citation were much more stable in that key papers continued to be built on year after year within the same topic. Few cocitation topics lasted for more than a few years, while a large fraction of direct citation topics lasted for ten years or more.

Given that we were very interested in the identification of emerging topics, we thus sought to combine the inherent stability of direct citation topics with the inherent instability of cocitation topics. We reasoned that if the same papers appeared in the first year or two of topics created using both methods, this would constitute two votes that these papers were part of an emerging topic. Using 15-year cocitation and direct citation models created from the same set of papers, we identified the top 25 emerging topics each year for four years, characterized those topics, and showed that these topics (and their key researchers) were associated with awards and honors at a much higher than expected rate [8.10]. Although this did not prove that these were the most emergent topics, it did suggest that the methodology was sound and identified a very reasonable set of emerging topics.

One thing that was not clear from this study was whether both models (direct citation and cocitation) were needed to identify emerging topics, or if direct citation alone would have worked just as well. It was clear that cocitation would not have worked on its own because the majority of new cocitation topics did not last for more than a year or two. The fact that nearly all of the top 25 emerging topics nominated by the combined approach were in the top 50 using only direct citation suggests that using a direct citation model alone would be a sufficiently accurate way of identifying emerging topics. We ultimately decided to start using direct citation with VOS/SLM clustering rather than using cocitation because:

1. The clusters were sufficiently accurate
2. The entire database could be clustered in a single calculation rather than linking annual models
3. It nominated a very defensible set of emerging topics; and
4. It requires far less computing time than using cocitation because there are far fewer links.

One known disadvantage of direct citation with respect to other document relatedness approaches is that it has tended to provide somewhat lower coverage. The major reason for this is that, in most calculations, the direct citation links used for clustering have been limited to those within a document set. This is particularly problematic for older documents. If, for instance, a ten-year dataset from 2001 to 2010 is used for clustering, some of the documents published in 2001 have no references within the dataset, and of these, documents not cited later are not linked and are thus not included in the clustering. Having made the decision to use direct citation clustering, and knowing of this disadvantage, we decided to investigate the effect of using what we call extended direct citation (EDC), which includes cited documents that are not in the dataset (i. e., nonsource documents) in the clustering. Using a 16-year dataset from Scopus, we compared a clustering using only source documents with one that included nonsource documents [8.51]. Inclusion of nonsource documents in the clustering significantly increased the coverage of source documents, from 85.6 to 96.4%. In addition to the 21.4 million source documents in the larger calculation, 22.0 million nonsource documents were also included in the clusters. While these nonsource documents have not been used for research evaluation, they can be used for other purposes. For instance, one can easily see the effect of books on topics in science, and one can also see how well the citation databases cover different subjects. In this study, it was shown that while source items comprise 88% of the documents in medical fields (in the 16-year source period), they only comprise 53% of documents in the social sciences. The social sciences could thus benefit much more than medicine from the additional coverage provided by nonsource documents.

Figure 8.5 graphically illustrates the difference between the direct citation (DC) and EDC approaches. EDC includes nonsource documents along with the links from source documents to nonsource documents, and enriches the clustering as a result. For instance, document N in Fig. 8.5 cannot be clustered using DC, but is included in a cluster using EDC. EDC not only increases coverage, but also improves the clustering of source documents because it uses more signal–it nearly doubles the number of edges included in the calculation, all of which provide additional information for clustering of the source documents. For instance, Fig. 8.5 shows document C being clustered with documents D and O using direct citation. However, when using the extended methodology, document C is no longer clustered with documents D and O because its links to nonsource documents pull it into another cluster. Recently, *Waltman* et al. [8.59] have shown experimentally that EDC can create more accurate clusters than DC, CC, or BC using a dataset of 273 000 documents in condensed-matter physics.

The accuracy of this EDC approach was also recently demonstrated in a study of a very large dataset. Although it was not compared with DC, EDC clearly outperformed BC and CC at multiple levels of granularity [8.52]. The most detailed EDC model in this study was created from 48.4 million documents with 582 million edges, resulting in a solution with 91 726 topics. This model will be described in more detail later along with several applications of the model that have policy implications.

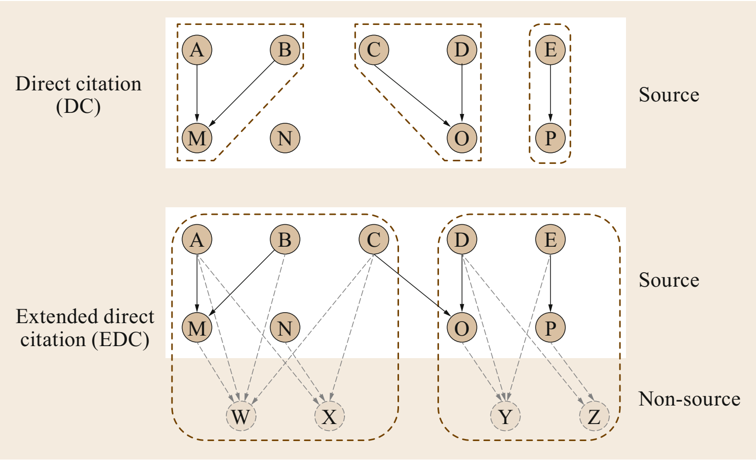


Fig. 8.5 Conceptual comparison of direct citation (DC) and extended direct citation (EDC) approaches. *Dashed boxes* indicate clusters

To this point, we have discussed only those studies of large-scale bibliometric networks published by SciTech Strategies and Centre for Science and Technology Studies at Leiden University (CWTS), being the two laboratories that have historically worked with such networks. Other laboratories with access to full databases have also started doing analysis of large-scale networks, although they have not created topic-level structures. Foster, Evans, and colleagues have been working with millions of records from Medline. Rather than clustering documents to create topics, they have been clustering chemicals [8.53] or authors and MeSH terms (chemicals, diseases, and methods) [8.54] for the purpose of characterizing the novelty of the linkages between these clusters. Although the numbers of clusters have not been large, these studies are noteworthy in that they are among the first studies using large-scale bibliometric data to explore edges rather than nodes. In addition, these studies have shown that the new or novel links that are formed each year are between objects that are already neighbors. Reference [8.54] in particular shows that most new connections come through things of a different type. For example, new connections between methods may come through diseases, and vice versa. Studies such as these are rich in that they are starting to explore and cross-link the Latourian feature space shown in Fig. 8.1.

As mentioned above, Microsoft Academic Graph (MAG) is a bibliographic database that has recently become available for research. These data were specifically used for the Web Search & Data Mining (WSDM) Cup 2016 where the challenge goal was to provide the best ranking of search results (i. e., other papers) using publications as queries. Teams from many countries participated in the challenge. Although they did not cluster the data into *topics*, the three teams with the highest cup scores all used network information to create their rankings. In other words, they all created document-to-document similarities using the network graph. These similarities could have easily been used to cluster the data, although this step was not taken because it was not part of the challenge. The methods to create similarities that were used included combined information from citations, authors and venues [8.60], an eigenfactor ranking method based on direct citation [8.55], and a modified relative citation ratio method based on cocitation clustering [8.19]. The direct citation eigenfactor solution is notable in that it has been expanded into a larger-scale recommendation engine combining inputs from MAG and several other databases [8.56], and has been converted into a web-based service that is publicly available.

The types of methods and applications described here for large-scale networks of scientific literature could be also applied to patents. However, we are unaware of any published work that clusters an entire patent database into patent topics. Thus, we present here a map of 2.8 million US patents from 1991 to 2011 to show the feasibility of such an approach. Patent citations were used as the direct citation links with the VOS code, resulting in a set of 27 118 clusters. A visual map of the clusters was created using cluster-cluster similarity calculated using the BM25 text relatedness measure, and is shown in Fig. 8.6. Labels were added to the map manually by looking at cluster contents in local areas of the map. Clusters were colored using the same colors used in our science maps (e. g., Fig. 8.4). This was done by assigning a color to each IPC4 code using the patent-to-paper links from author-inventors [8.61], and then determining the dominant color for each cluster based on its IPC4 codes.

As shown in Fig. 8.6, many patent clusters link to science, particularly computer science and electrical engineering (pink), other engineering (cyan), chemistry

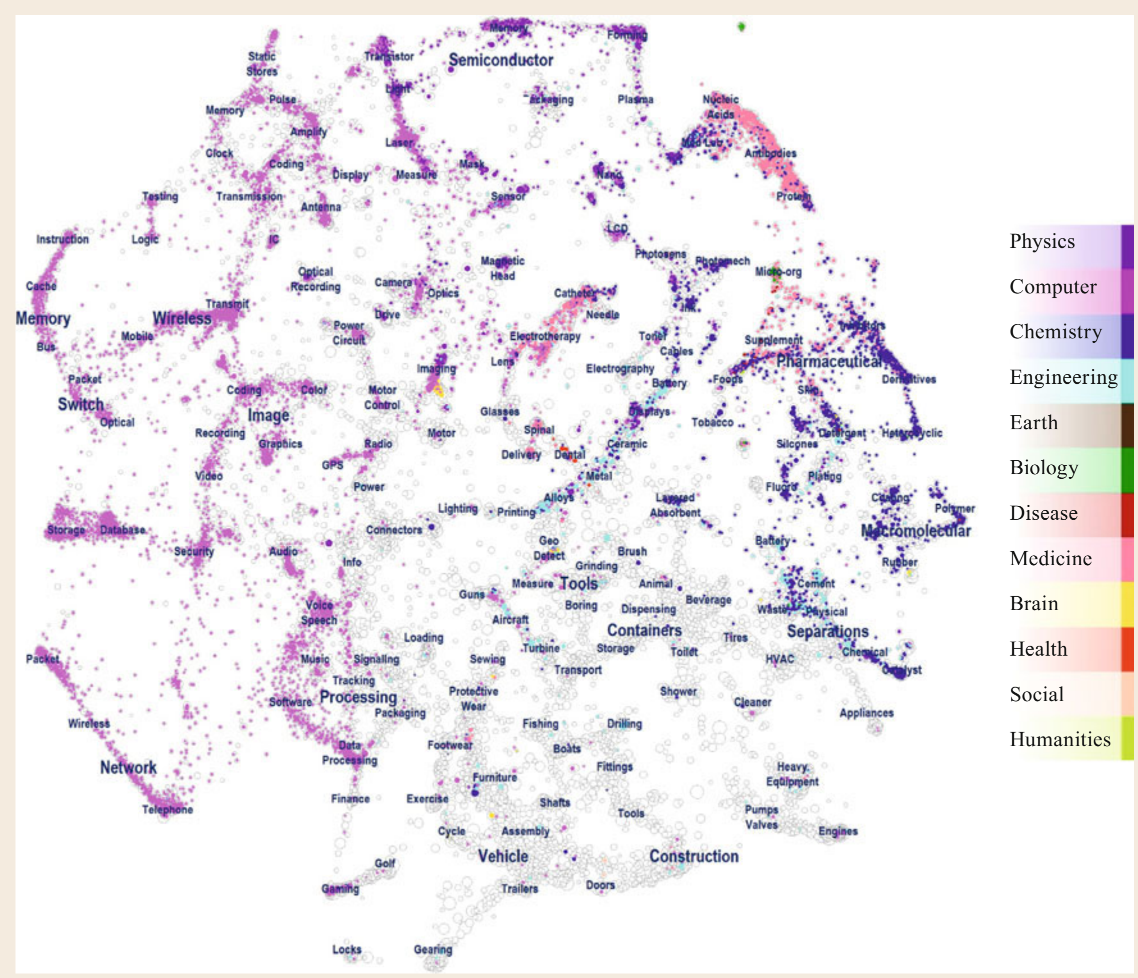


Fig. 8.6 Visual map of clusters of patents based on direct citation where each *dot* represents a cluster

(blue), and medicine (salmon). However, many patent clusters do not have strong links to science. These have been colored gray in the map, and most such clusters are focused on consumer goods. As with the topic-level clusterings of the scientific literature mentioned above, patent clusters can be used for many applications, such as portfolio analysis and the identification of emerging clusters.

8.3.2 Networks from the FUSE Program

In 2011, the US Intelligence Advanced Research Projects Activity (IARPA) funded the Foresight and Understanding from Scientific Exposition (FUSE) program with the intent of developing automated methods to identify technical emergence from literature sources. Four research teams were chosen to participate in this effort. As with many programs funded by US intelligence agencies, relatively little information about FUSE and its outcomes is publicly available. It can be deduced from published outputs, however, that very large networks were created and used in the task. For instance, a short program review in Nature mentions that FUSE mined "millions of papers and patents in both English and Chinese" [8.62]. The article also mentions that FUSE "has identified several hundred indicators, such as new collaborations or expressions of excitement in text, that highlight emerging areas." Work from the team headed by researchers at Columbia University makes it clear that network properties from 48 million WoS articles based on direct citation links and coauthorship links were among those indicators [8.63]. The combined cocitation and direct citation network analysis to identify emerging topics mentioned above [8.10] also came from the FUSE program; large-scale network analysis was routinely used in the FUSE program, although very few published outputs mention this fact. Finally, FUSE technology is also being commercialized by Meta, who licensed the technology developed in the FUSE program by SRI International to mine very large literature databases. Meta has recently joined the Chan Zuckerberg initiative with the intention of making their technology available to people worldwide to help them understand the state of scientific knowledge (https://meta.com). Although detailed information about FUSE outputs is sparse, we mention it here to illustrate that large-scale network analyses are perhaps more advanced than those in the academic world may realize.

8.3.3 Author Disambiguation

Author disambiguation is another application of large-scale bibliometric networks. Although most author disambiguation studies have been done with small datasets from a single field of science, there have been several studies that have developed methods to disambiguate author names over full large-scale databases. Common among the five studies listed in Table 8.3 are that they were each done using a full large-scale database, and each clustered author-paper pairs to identify unique authors. The first of these efforts, the Author-ity model created by *Torvik* and *Smalheiser* [8.64], used 15.3 million papers from Medline. It used multiple features including title words, coauthors, MeSH terms, affiliations, and email addresses to estimate the probability that two papers sharing the same author name (last name with first initial) were actually written by the same person. Each author-paper pair was assigned to one of 6.7 million individuals. The Author-ity model was later adapted for use on the US patent database by *Li* et al. [8.65], who included secondary data sources such as geospatial country files and geographic name unification files as inputs to their clustering. Improved author and inventor assignments enable more accurate characterization of these people and their mobility.

Liu et al. revisited the disambiguation of biomedical authors by working with 22 million records from PubMed [8.66]. They attributed their improvement over the Author-ity methodology to their "machine-learning method driven by a large-scale training set and the clustering algorithm regulated by a name compatibility scheme preferring precision." Their disambiguation results have been used by PubMed since May 2012 to provide better search results in those 36% of queries that contain author names. *Schulz* et al. designed a disambiguation method that is fundamentally different that those already listed in that it is based primarily on shared references and citations [8.67]. The assumption behind this approach is that two papers written by the same author are much more likely to be connected by a citation link (i. e., a self-citation) than two papers written by two different authors sharing the same name. This approach was applied to the entire WoS database (47 million articles) with good results. *Caron* and *van Eck* [8.68] applied a similar approach to the full WoS database, using author and article-based features along with citation characteristics. Author disambiguation is an extremely important application of large-scale bibliometric networks because lumping together of multiple authors has a relatively large (unwanted and negative) effect on network measures [8.69], and can lead to incorrect analysis.

8.3.4 Other Relevant Analyses

Sections 8.3.1–8.3.3 have mentioned those studies that have created and analyzed large-scale bibliometric networks where the network has been partitioned into clusters or has been considered as a whole. In this section, we consider several studies of two other types: a) those that analyze individual links in large-scale data, but that do not consider the entire network; and b) studies that are too small to be considered large-scale (using our criteria), but that use the data sources from Sect. 8.2.2 in important ways that could be scaled to larger-scale analysis.

Studies of Individual Links

Citation analysis has been used for many years to try to predict which papers will end up being the most highly cited. Recent studies using millions of articles from WoS and Scopus have characterized individual cocitation links by their novelty, and have correlated the presence of novel or interdisciplinary cocitation links with scientific impact. *Uzzi* et al. used cocited journal-

Table 8.3 Large-scale author disambiguation studies based on the clustering of author–paper pairs to identify unique authors

Study	Data	Unit	Sim	#Nodes
Torvik and *Smalheiser* [8.64]	Medline	Authors	Co-occur	6.7 M
Li et al. [8.65]	USPTO	Authors	Co-occur	2.67 M
Liu et al. [8.66]	PubMed	Authors	Co-occur	10.2 M
Schulz et al. [8.67]	WoS	Authors	Co-occur	6.5 M, h>1
Caron and *van Eck* [8.68]	WoS	Authors	Co-occur	All WoS

journal relationships to determine whether any pair of cited references is typical or atypical, and found that the highest impact articles are likely to have a combination of typical and atypical cocited journal pairs [8.70]. *Boyack* and *Klavans* reproduced their study and showed that the disciplinary effects were far more prevalent than claimed by Uzzi et al. [8.71]. Multidisciplinary journals like Science, Nature, and PNAS were disproportionately involved in atypical combinations. Regardless of the details, both studies confirmed that highly cited articles are, in general, enriched in atypical (or novel) combinations of cocited references. A similar study by *Larivière* et al. [8.72] used cocited journal categories rather than cocited journals, and showed that interdisciplinary research, as measured by cocited journal categories, leads to higher-impact research.

Additional Studies of Key Data Sources

In Sections 8.2.1 and 8.2.2 we made the point that the types of data sources that we analyze should be expanded well beyond just the scientific literature to gain a better understanding of the science system. We thus review here a few studies that are not large-scale network studies to show the state of the art regarding some of these additional data sources. Although we showed a large-scale patent map in Sect. 8.3.1, most patent studies have dealt with smaller sets and have not clustered individual patents. For example, *Kay* et al. [8.73] created an overlay patent map of 400 International Patent Classification (IPC) categories from a set of 760 000 European patents using cocategory analysis. The map is useful for showing portfolio overlays, but lacks the topic-level details.

Two recent studies have created models from grant data using a latent Dirichlet allocation (LDA) topic modeling approach. *Talley* et al. used topic modeling to calculate grant-grant similarities on a set of 80 000 NIH grants, and then generated a visual map of the grants using DrL [8.74]. Nichols used topic modeling to assign 170 000 NSF grants to topics, and then characterized each grant in terms of its interdisciplinarity (i. e., its profile over multiple topics). We wish to make clear that topic modeling is not network analysis. It is a statistical technique based on text that probabilistically assigns documents to topics. However, topic modeling can be used to create topic structures than can then be subjected to network analysis. These two studies, while not large-scale network analyses, are the largest such efforts using grant data of which we are aware, and thus are worthy of consideration. It is our hope that as more grant data become available, along with more data linking grants to their outputs, large-scale analysis of these data will become more common. Along those lines, a recent poster presentation by *Freyman* et al. [8.20] reports on work that has been done to create linkage data between scientific articles, patents, grants, and technology licensing agreements by matching records across multiple databases. They have identified more than 400 000 links from articles to NSF grants, more than five million links from US patents to articles, and more than 10 000 links from agreements to patents. This type of linking has obvious policy implications, and we suggest that much more of this type of work needs to be done in the future.

Finally, as mentioned above, the full text of scientific articles is becoming more available with time. The FUSE program was very involved with mining full-text articles for features such as named entities, citation sentiment, and the rhetorical status of sentences [8.63]. Simpler features, such as the number and location of in-text references, have the potential to impact large-scale bibliometric networks. First steps have recently been taken to understand the distributions of references in full text [8.75, 76]. Reference counts, ages, and positions follow certain patterns, with some field dependency. These data have the potential to increase the accuracy of similarity measures, which would lead to more accurate identification of topics, and perhaps even to better measures of impact. Work to explore these possibilities is planned.

8.4 The STS Global Model of Science

Having reviewed relevant literature on the creation and analysis of large-scale bibliometric networks, we now present the current SciTech Strategies (STS) model of science in detail, and show how it can be used for various applications. As mentioned above, this model was recently shown to create topics that were very consistent with the referencing patterns of experts (those writing papers with at least 100 references) [8.52], and we thus use it as an example of a reasonably accurate model of the structure and dynamics of science. We do not suggest that this is the best model possible, but rather that it is reasonably comprehensive and accurate, and that it has been found to be useful for several different types of analysis by clients. Other topic-level models of science that are equally (or perhaps even more) accurate and useful can certainly be created.

The original STS model of science consists of 48 398 815 documents from Scopus. Of these, 24 615 844 documents are indexed source documents from Scopus 1996–2012, while the remaining 23 782 971 are nonsource documents that were each cited at least twice by the set of source documents. Including nonsource documents extends the coverage of the model to include important science not indexed by Scopus, including many books. In particular, the social sciences are heavily augmented when nonsource documents are included [8.51]. The set of 48.4 million documents are connected by 582 million direct citation links, which were used to create the model. Clustering was done using the SLM approach from CWTS at Leiden University [8.33] that has recently been shown to be among the most accurate clustering algorithms available [8.77]. The similarity value (a_{ij}) between each pair of papers i and j was set to $1/k$ where k is the number of edges (both citing and cited) for paper j [8.25]. Note that symmetry is not assumed; a_{ij} and a_{ji} are typically different since the numbers of edges for papers i and j are rarely the same. The CWTS methodology allows a desired minimum cluster size and resolution value to be specified–these can be tuned to produce a desired number of clusters. We desired a solution with approximately 100 000 clusters. Our experience is that at the 100k cluster level: a) experts can easily differentiate between clusters [8.58]; and b) funding data can be assigned to topics and is highly correlated with topic-level metrics [8.78]. Using a minimum cluster size of 50 papers and a resolution of 3×10^{-5}, a cluster solution was obtained with 91 726 clusters above the minimum size. Each cluster represents a topic, and is comprised of the papers on that topic and the community of researchers working on that topic. A total of 134 066 (0.28%) of the documents ended up in clusters with fewer than 50 documents. These clusters are small and disconnected from the rest of the graph, and thus are not considered further. Creation of a model such as this obviously entails many choices in terms of relatedness measures, clustering algorithms, thresholds, etc. The choices we have made along with the reasoning for those choices is available in [8.78] and is not reproduced here.

Figure 8.4 shows a map of the 91 726 topics created to provide a visual depiction of the structure of science. It was created using the following process:

1. The Similarity between pairs of topics was calculated from the titles and abstracts of the documents in each topic using the BM25 similarity measure.
2. The resulting similarity list was filtered to keep only the top-n (between 5 and 15) similarities per topic.
3. A layout of the topics was created using the DrL algorithm [8.41], which gives each topic an x,y position based on the similarity graph.

One might wonder why an additional direct citation step is not used to create the visual map. This could certainly be done, but we have found that using text creates a more accurate and visually appealing map than using a citation-based measure for this step [8.46]. Each of the 91 726 topics in the map has been designated as belonging primarily to one of twelve high-level fields, and is correspondingly colored in Fig. 8.4. These designations were made by assigning papers to fields using the UCSD journal-to-field assignments [8.47].

At a high level, the field structure shown in Fig. 8.4 is very similar to that of many other global science maps, including the consensus map of science [8.2]. Physics, chemistry, and engineering are highly related, and are adjacent to each other. The medical areas (disease, medicine, health sciences, brain sciences) are also adjacent to each other. Biology is adjacent to chemistry and medicine, earth sciences is primarily adjacent to engineering, and social sciences are adjacent to health sciences, while computer science lies between physics (which includes mathematics) and the social sciences.

This model has a number of features that attest to its robustness and that make it useful for a variety of applications. In terms of robustness, nearly 90% of the source documents in the map are found in their dominant cluster. For example, for the nearly four million documents published in 2009 and 2010, 89.2% of them are in the cluster to which they have the greatest number of links (combined citing and cited), and on average 58% of their links are to that dominant cluster. There is also a relatively high correlation between text and links in the model. Although created using direct citation, the topics are very well differentiated from each other textually. Of papers with a sufficiently large abstract, 75% can be accurately assigned to the cluster in which they reside using the BM25 similarity measure based on title and abstract [8.78]. This high level of topic differentiation, both in terms of citation and text, enables relatively accurate linking of documents of other types to individual topics.

The model also exhibits size distributions that are consistent with what we would expect given other complex networks. Figure 8.7 shows distributions for topic sizes (numbers of documents) along with the numbers of unique authors per topic for 2010. These are compared to the distribution of US city sizes. All curves are roughly linear on a log–log scale at the upper end of the distribution, and then each tails off at the lower end. The slope for topics is lower for all years combined than for a single year; combining years smooths the distribution. In addition, the fact that some topics are very large, and that others are very small, correlates with *Derek de Solla Price*'s notions of Big Science and Little Science [8.79]. Big Science refers to ma-

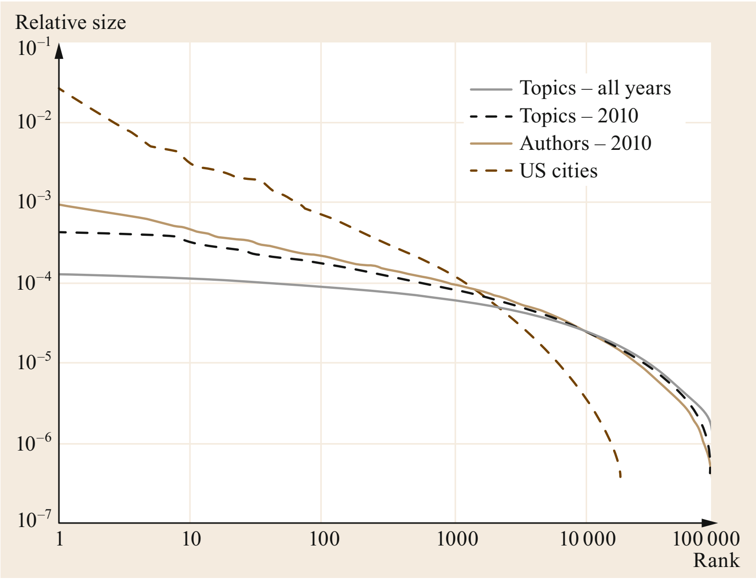


Fig. 8.7 Size distributions for topics and authors from the STS model of science and US city populations

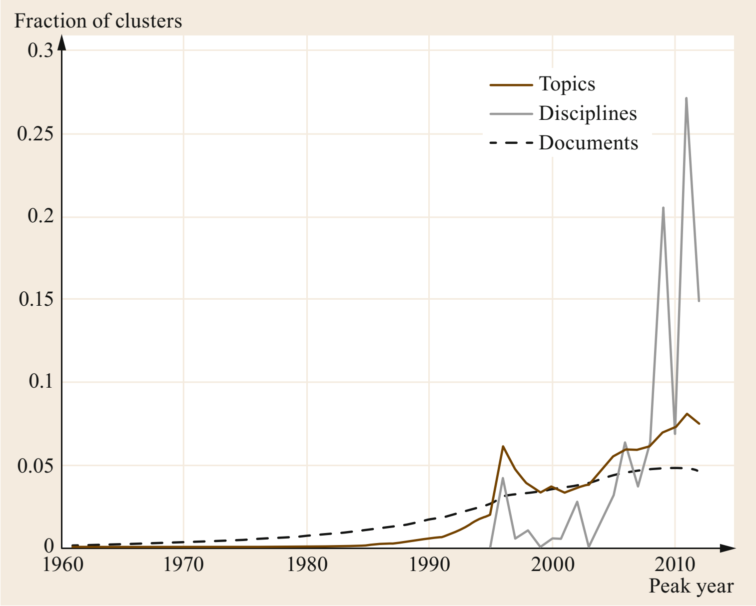


Fig. 8.8 Distributions of peak years for different partitioning strategies. *Spikes* in 1996 are due to boundary effects in the Scopus database–references were only available for papers starting in 1996

jor investments in infrastructure and the corresponding creation of research communities involving potentially hundreds of researchers. Big Science communities are especially common in biomedical research and high-energy physics, and are occasionally found in other fields as well. Significant funding for infrastructure is a necessary prerequisite for Big Science. Little Science can survive on the existing infrastructure and requires much less external funding.

The evolution of scientific areas, and more particularly the identification of emerging areas, is a topic that is of interest to most people involved in the science system. Identification of emerging areas is highly dependent upon the granularity of a model. Figure 8.8 shows the fractions of clusters that peak by year for our topic-level model, and also for an aggregation of topics to 188 discipline-level clusters. These distributions are both compared to a curve showing the fraction of documents by year in the model; one might expect the distribution of peak years to be similar to the overall distribution of documents given that topics in science ebb and flow with the changing priorities of agencies, researchers, etc. Well over half (69%) of disciplines peak in one of the most recent four years. To a large degree,

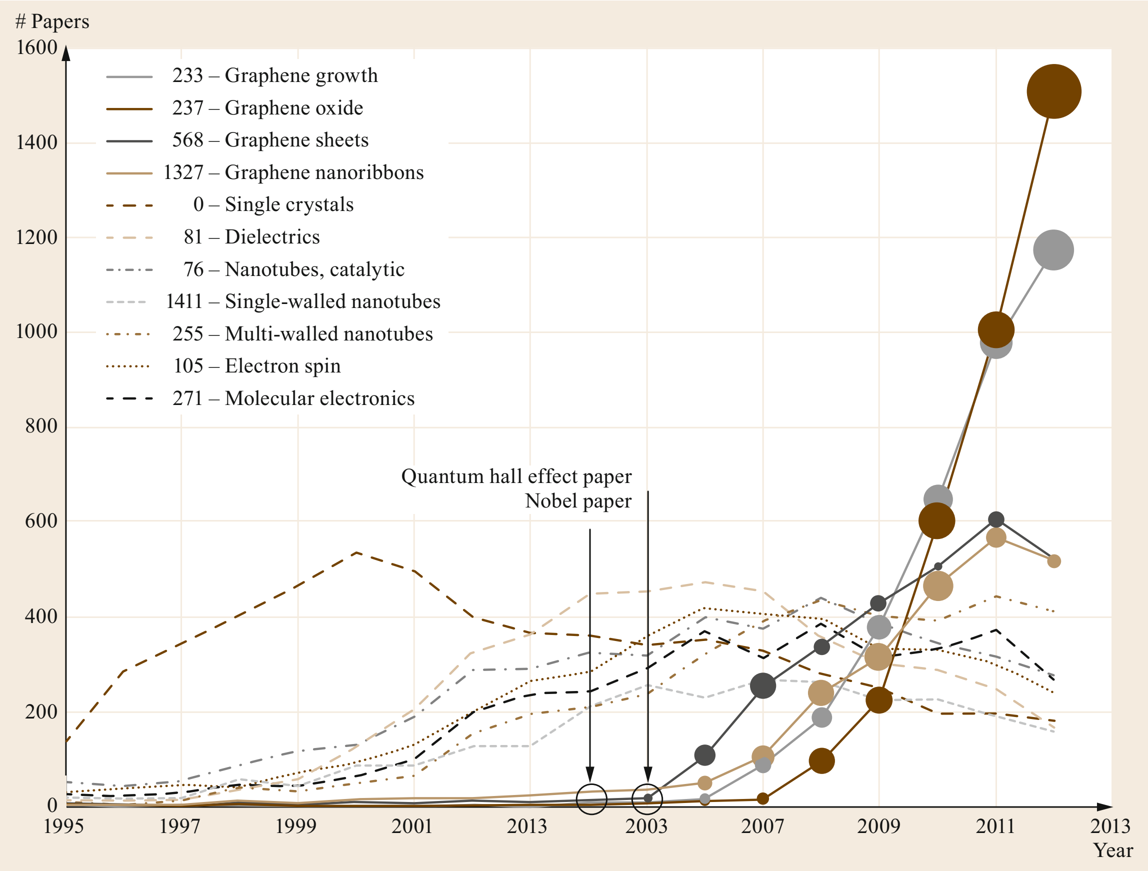


Fig. 8.9 Detailed histories for emerging graphene-related topics and other topics from which graphene researchers migrated. *Dots* represent the numbers of hot papers (top 1%) in the four graphene topics by year

disciplines grow monotonically, mirroring the growth of the entire model, and thus very few disciplines are currently in decline. In contrast, topics have peak years that are more evenly distributed. Topic peak years are far more similar to the temporal document distribution than are discipline peak years. This distinction is important for the identification of emerging topics because, at the discipline level, most scientific areas look like they are emerging or have recently emerged. Topic-level clusters provide a much more differentiated view of evolution and of scientific areas, and are far better than discipline-level clusters at identifying emerging areas.

Topic-level clusters also enable detailed analysis of emerging topics and their relationships with existing and/or declining topics. Graphene is an example of a scientific subject with multiple emerging topics [8.10]. Figure 8.9 shows four distinct graphene-related topics that emerged in the mid-to-late 2000s, all of which experienced extremely rapid growth at that time. None of these four topics sprang out of thin air; each had a low level of activity in the late 1990s and early 2000s, and all four topics at that time were focused on graphite (another carbon structure) research until the (Nobel Prize-winning) breakthrough of graphene paper was published in 2004. Once the breakthrough was made, and over the next several years, large numbers of researchers shifted their research to graphene-related topics, migrating from existing topics, many of which started to decline as research on graphene emerged and grew. We note that the topics that supplied the largest numbers of graphene researchers were inherently related to graphene, and included research on carbon nanotubes, single crystals, and electronic properties. It was thus natural for researchers in these topics to migrate to more attractive (emergent) research topics.

Another feature of our topic-level model is that it shows that science is inherently multidisciplinary in many areas. Figure 8.10 shows a zoomed-in region of the map from Fig. 8.4. Although there are some areas of the map where only a single color (see the Fig. 8.4 legend) is found, there are also many regions of the map where topics with four, five, or six different colors–indicating research from several different

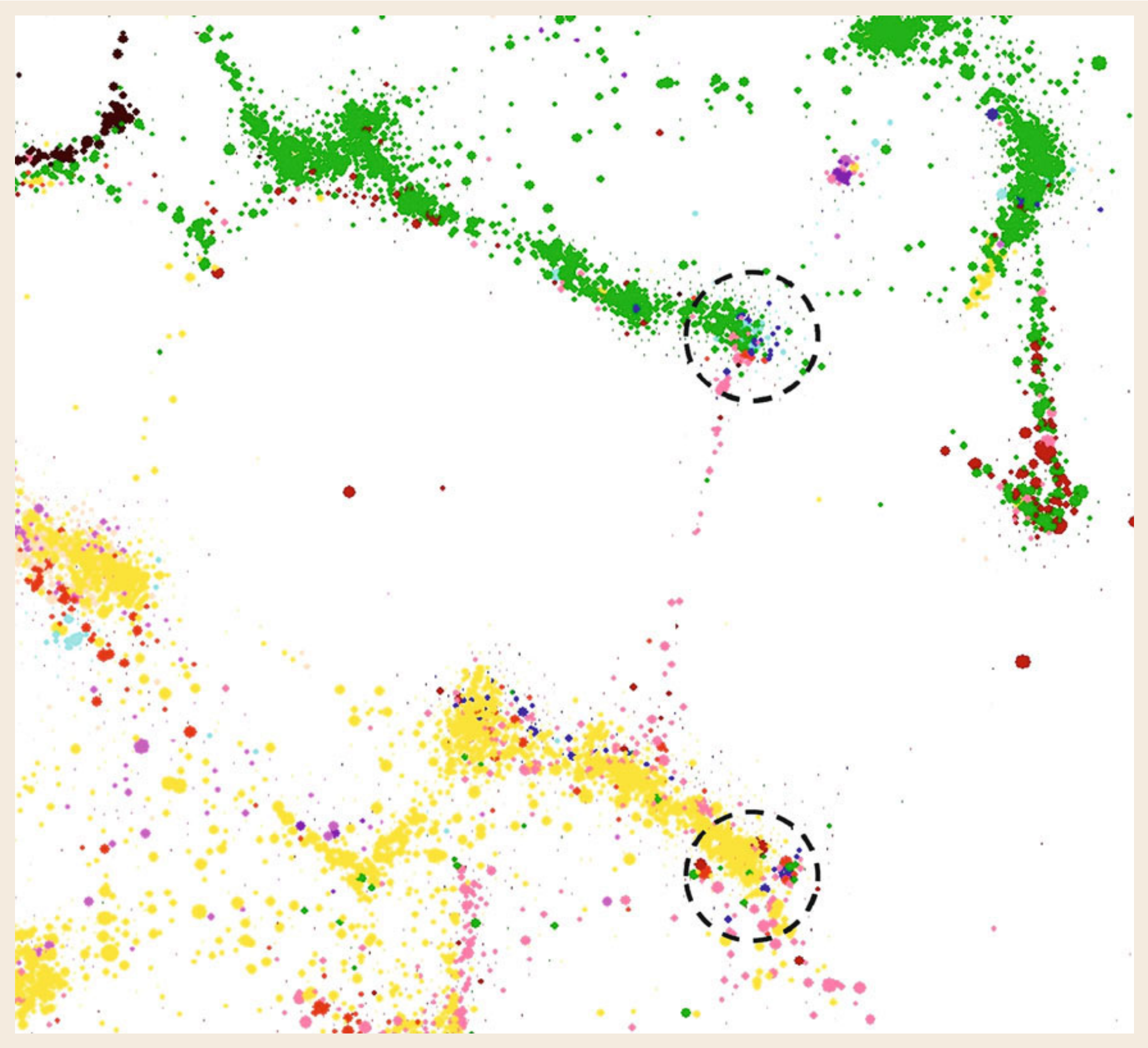


Fig. 8.10 Detail from the map of Fig. 8.4. Local scientific areas with adjacent topics of multiple colors are inherently multidisciplinary

fields–are adjacent to one another. For example, the circled region near the top of the map contains clusters of six colors–this region contains topics from biology, medicine, health sciences, engineering, chemistry, and computer science all in close proximity. The circled region near the bottom of the map contains adjacent topics from brain sciences, medicine, biology, chemistry, and computer science. Of course, there are also areas of the map dominated by a single field (color). Recall that this map was created using a sequentially hybrid process. Clusters (or topics) are groups of papers related through direct citation, while the positions of the clusters are based on textual similarity between clusters. Thus, neighboring clusters of different colors are an indication that very similar language is used in multiple disciplines, and thus that an area of science with many colors is multidisciplinary. Current science contains a mix of disciplinary and multidisciplinary work. However, given that there are many areas in science receiving contributions from several different broad fields, it raises the question in our minds as to what *multidisciplinarity* really means today.

The STS model of science has also been used in several other studies with practical applications. For example, it has been used as a template to create a visual map of PLoS thesaurus terms that can then be used as an overlay map in its own right [8.80]. Its topics have been aggregated into higher-level groupings that have then been used to characterize the research focuses of different nations [8.81].

Perhaps most importantly from the perspective of topics, StarMetrics funding data have been linked to individual topics. Given that papers can be assigned to topics with 75% accuracy, and since project abstracts are similar in content to abstracts in scientific papers, it is reasonable to assume that projects can be accurately assigned to topics with similar accuracy. Each project with sufficient text has thus been individually linked to those topics to which it is most similar, which enables topics to be differentiated by the amount of funding they receive. We find that the most highly prominent topics–those that are most visible–receive about \$90 000 annually per US author from StarMetrics funding sources. Topics that are the least prominent receive only about \$2000 per US author annually from those same funding sources. Furthermore, we find that these data enable prediction of those topics that will receive increased funding in the future, which is very important information for a number of stakeholders in the science system [8.78]. Although this analysis only accounts for a small amount of the worldwide funding, it is nevertheless a substantial amount ($\approx$ \$160 billion) and thus shows the potential for topic-level metrics based on funding.

8.5 Summary and Implications

The analyses presented in Sect. 8.4 suggests that a robust and accurate topic-level model of science can be created using comprehensive citation data and network techniques, and that models such as this can be usefully applied to pertinent questions about the science system. Although we use direct citation links, the CWTS clustering algorithms to create our models, and prefer models with around 100 000 topics, it is clear that very useful models can also be created using other citation or text-based approaches with other algorithms, and at different levels of granularity. Although we prefer to use full databases, and have evidence that topics in a global model are more accurate than topics in a local model, we also acknowledge that topics created from smaller datasets can be extremely useful to answer practical questions, provided that the data are well matched to the question. The methods presented in this chapter are not specific to large-scale models using full databases, but can be used with datasets of any size.

Regarding the future, we suggest that Latour's model represents an opportunity. Network research related to research communities and the science system has traditionally been done where data has been most readily available–i. e., using papers, journals, authors, and patents. Much less work has been done using data on funding, equipment, and the types of laboratory artifacts that are often the building blocks of scientific work (e. g., chemicals, models, etc.). Still less work has been done with bipartite or multipartite networks comprised of many of these features. Modeling of single features (e. g., papers) has helped us to gain a rudimentary understanding of the science system. However, given that science is comprised of so many features with their myriad connections, it is likely that a clearer understanding of the science system can only be gained through analysis of networks comprised of multiple features.

Several directions for future work seem particularly promising to us. One could focus on the nature and fate of transient authors. What do they do? Are they more likely to be members of corporations? If their jobs are at risk, we can hypothesize that transient authors that publish when there is a spike are more likely to change employment over the next few years. Transient authors that publish during normal times (a much lower percentage of transient authors) are less likely to change employment over the next few years. Overall, a better understanding of transient authors is needed. We simply don't know much about the very large author population.

From a pragmatic perspective, any indicator that can predict when a topic becomes 'unattractive' is extremely useful. The current emphasis is on indicators of attractiveness (hot papers, high vitality, and citation impact). There is corresponding emphasis on (positive) turning points (a sudden shift to becoming attractive). There is far less effort on developing indicators that a topic is switching from being attractive to being unattractive. These indicators would be especially useful if they can predict negative changes in funding.

From a science policy perspective, the role of instrumentation is an underresearched, but extremely important, area to focus on. Kuhn, Price and Latour have all emphasized the importance of instrumentation, and the role of instrumentation has played a central role in the development of the model presented in this paper. But it has been our experience, and it is our strong suspicion, that future developments in instrumentation are far more likely to come from software instead of hardware. For example, in 1985 we needed to rent a mainframe computer and create our own algorithms to cluster 100 000 documents at a cost that exceeded \$10 000 / year [8.82]. Computing power today is ubiquitous. If we do not own computers capable of large calculations, cloud computing is widely available at low cost. Open-source algorithms such as SLM that cluster 50 million documents are also available. Highly accurate and low-cost experimentation using bibliometric data is possible today. We do not think this is an isolated event. Computing capability continues to increase. Advances in software, particularly open-source software, are being made perhaps even more rapidly. If this trend continues, the infrastructure requirements of large topics may lessen, and the long-term shift towards Big Science may be reversed. Eventually, we will realize that instrumentation is fundamentally changing because of the information revolution, and take this into account in science policy studies.

Finally, from a theoretical perspective, we find that it is extremely short-sighted to exclude the motives of the circle of actors in Latour's model from our analysis. Price placed great emphasis on the motives of the pure scientist–the 'search for knowledge' that should be protected from the practical requirement of solving social problems. Latour emphasized the need to find common motives (e. g., the search for a cure of a disease by funders is aligned with the potential discoveries from doing genetic research) and referenced some of the broad motives normally associated with R&D investment (e. g., health, national security, and economic welfare). Instead of arguing which motives take priority, we strongly suspect that common motives can be found by looking at the full set of altruistic motives that are expressed by people [8.83]. This is a line of investigation that is almost completely unexplored. We strongly suspect that by understanding the motives of stakeholders, we will gain a much more thorough understanding of the science system, and would love to see others join us in pursuing research along these lines.

References

8.1 H.F. Moed, W. Glänzel, U. Schmoch (Eds.): *Handbook of Quantitative Science and Technology Research* (Springer, Dordrecht 2004)

8.2 R. Klavans, K.W. Boyack: Toward a consensus map of science, J. Am. Soc. Inf. Sci. Technol. **60**(3), 455–476 (2009)

8.3 I. Rafols, A.L. Porter, L. Leydesdorff: Science overlay maps: A new tool for research policy and library management, J. Am. Soc. Inf. Sci. Technol. **61**(9), 1871–1887 (2010)

8.4 T. Velden, K.W. Boyack, J. Gläser, R. Koopman, A. Scharnhorst, S. Wang: Comparison of topic extraction approaches and their results, Scientometrics **111**, 1169–1221 (2017)

8.5 E. Garfield: Citation indexes for science: A new dimension in documentation through association of ideas, Science **122**, 108–111 (1955)

8.6 E. Garfield, I.H. Sher, R.J. Torpie: *The Use of Citation Data in Writing the History of Science* (Institute for Scientific Information, Philadelphia 1964)

8.7 H. Small: Co-citation in the scientific literature: A new measure of the relationship between two documents, J. Am. Soc. Inf. Sci. **24**, 265–269 (1973)

8.8 H. Small, B.C. Griffith: The structure of scientific literatures, I: Identifying and graphing specialties, Soc. Stud. Sci. **4**, 17–40 (1974)

8.9 Clarivate Analytics: Research Fronts 2016, https://clarivate.com/wp-content/uploads/2017/10/Research_Fronts_2016_Report_EN.pdf (2016)

8.10 H. Small, K.W. Boyack, R. Klavans: Identifying emerging topics in science and technology, Res. Policy **43**, 1450–1467 (2014)

8.11 D. Rotolo, D. Hicks, B. Martin: What is an emerging technology?, Res. Policy **44**(10), 1827–1843 (2015)

8.12 T.S. Kuhn: *The Structure of Scientific Revolutions*, 2nd edn. (Univ. Chicago Press, Chicago 1970)

8.13 N.C. Mullins: *Theories and Theory Groups in Contemporary American Sociology* (Harper Row, New York 1973)

8.14 D. Crane: *Invisible Colleges. Diffusion of Knowledge in Scientific Communities* (Univ. Chicago Press, Chicago 1972)

8.15 K.W. Boyack: Investigating the effect of global data on topic detection, Scientometrics **111**, 999–1015 (2017)

8.16 R. Klavans, K.W. Boyack: Using global mapping to create more accurate document-level maps of research fields, J. Am. Soc. Inf. Sci. Technol. **62**(1), 1–18 (2011)

8.17 B. Latour: *Science in Action* (Harvard Univ. Press, Cambridge 1987)

8.18 D. Herrmannova, P. Knoth: An analysis of the Microsoft Academic Graph, D-Lib Magazine (2016), https://doi.org/10.1045/september2016-herrmannova

8.19 S. Ribas, A. Ueda, R.L.T. Santos, B. Ribeiro-Neto, N. Ziviani: UFMG/LATIN at WSDM Cup 2016: Simplified relative citation ratio for static paper ranking. In: *9th ACM International Conference on Web Search and Data Mining* (ACM, San Francisco 2016)

8.20 C. Freyman, J. Byrnes, T. Muezzinoglu: Knowledge flows: Linking big data sets. In: *OECD Blue Sky III* (In, Ghent 2016), http://de.slideshare.net/innovationoecd/110-freyman-knowledge-flows-linking-big-dataset

8.21 A. Breitzman, P. Thomas: The emerging clusters model: A tool for identifying emerging technologies across multiple patent systems, Res. Policy **44**, 195–205 (2015)

8.22 K. Börner, C. Chen, K.W. Boyack: Visualizing knowledge domains, Annu. Rev. Inf. Sci. Technol. **37**, 179–255 (2003)

8.23 H. Small: Update on science mapping: Creating large document spaces, Scientometrics **38**(2), 275–293 (1997)

8.24 K.W. Boyack, R. Klavans: Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accu-

rately?, J. Am. Soc. Inf. Sci. Technol. **61**(12), 2389–2404 (2010)
8.25 L. Waltman, N.J. van Eck: A new methodology for constructing a publication-level classification system of science, J. Am. Soc. Inf. Sci. Technol. **63**(12), 2378–2392 (2012)
8.26 N.J. van Eck, L. Waltman: How to normalize cooccurrence data? An analysis of some well-known similarity measures, J. Am. Soc. Inf. Sci. Technol. **60**(8), 1635–1651 (2009)
8.27 K.W. Boyack, D. Newman, R.J. Duhon, R. Klavans, M. Patek, J.R. Biberstine, B. Schijvenaars, A. Skupin, N. Ma, K. Börner: Clustering more than two million biomedical publications: Comparing the accuracies of nine text-based similarity approaches, PLoS One **6**(3), e18029 (2011)
8.28 S. Fortunato: Community detection in graphs, Phys. Rep. **486**(3–5), 75–174 (2010)
8.29 L. Šubelj, N.J. van Eck, L. Waltman: Clustering scientific publications based on citation relations: A systematic comparison of different methods, PLoS One **11**(4), e154404 (2016)
8.30 M. Rosvall, C.T. Bergstrom: Maps of random walks on complex networks reveal community structure, Proc. Natl. Acad. Sci. USA **105**(4), 1118–1123 (2008)
8.31 S. Bae, D. Halperin, J.D. West, M. Rosvall, B. Howe: Scalable and efficient flow-based community detection for large-scale graph analysis, ACM Trans. Knowl. Discov. Data **11**(3), 32 (2017)
8.32 V.D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre: Fast unfolding of communities in large networks, J. Stat. Mech. Theory Exp. **10**, P10008 (2008)
8.33 L. Waltman, N.J. van Eck: A smart local moving algorithm for large-scale modularity-based community detection, Eur. Phys. J. B **86**, 471 (2013)
8.34 C. Chen, F. Ibekwe-SanJuan, J. Hou: The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis, J. Am. Soc. Inf. Sci. Technol. **61**(7), 1386–1409 (2010)
8.35 C. Chen: *CiteSpace: A Practical Guide for Mapping Scientific Literature* (Nova Science, New York 2016)
8.36 N.J. van Eck, L. Waltman: Software survey: VOSviewer, a computer program for bibliometric mapping, Scientometrics **84**(2), 523–538 (2010)
8.37 M.J. Cobo, A.G. Lopez-Herrera, E. Herrera-Viedma, F. Herrera: Science mapping software tools: Review, analysis, and cooperative study among tools, J. Am. Soc. Inf. Sci. Technol. **62**(7), 1382–1402 (2011)
8.38 M.J. Cobo, A.G. Lopez-Herrera, E. Herrera-Viedma, F. Herrera: SciMAT: A new science mapping analysis software tool, J. Am. Soc. Inf. Sci. Technol. **63**(8), 1609–1630 (2012)
8.39 T. Kamada, S. Kawai: An algorithm for drawing general undirected graphs, Inf. Process. Lett. **31**, 7–15 (1988)
8.40 T.M.J. Fruchterman, E.M. Reingold: Graph drawing by force-directed placement, Softw. Pract. Exp. **21**(11), 1129–1164 (1991)
8.41 S. Martin, W.M. Brown, R. Klavans, K.W. Boyack: OpenOrd: An open-source toolbox for large graph layout, Proc. SPIE Int. Soc. Opt. Eng. **7868**, 786806 (2011)
8.42 A. Mrvar, V. Batagelj: Analysis and visualization of large networks with program package Pajek, Complex Adapt. Syst. Model. **4**, 6 (2016)
8.43 M. Bastian, S. Heymann, M. Jacomy: Gephi: An open source software for exploring and manipulating networks. In: *3rd Int. AAAI Conf. Weblogs Soc. Media* (2009)
8.44 K. Sparck Jones, S. Walker, S.E. Robertson: A probabilistic model of information retrieval: Development and comparative experiments. Part 1, Inf. Process. Manag. **36**(6), 779–808 (2000)
8.45 K. Sparck Jones, S. Walker, S.E. Robertson: A probabilistic model of information retrieval: Development and comparative experiments. Part 2, Inf. Process. Manag. **36**(6), 809–840 (2000)
8.46 K.W. Boyack, R. Klavans: Creation of a highly detailed, dynamic, global model and map of science, J. Assoc. Inf. Sci. Technol. **65**(4), 670–685 (2014)
8.47 K. Börner, R. Klavans, M. Patek, A.M. Zoss, J.R. Biberstine, R.P. Light, V. Lariviere, K.W. Boyack: Design and update of a classification system: The UCSD map of science, PLoS ONE **7**(7), e39464 (2012)
8.48 K.W. Boyack: Using detailed maps of science to identify potential collaborations, Scientometrics **79**(1), 27–44 (2009)
8.49 R. Klavans, K.W. Boyack: Toward an objective, reliable and accurate method for measuring research leadership, Scientometrics **82**(3), 539–553 (2000)
8.50 J. Ruiz-Castillo, L. Waltman: Field-normalized citation impact indicators using algorithmically constructed classification systems of science, J. Informetr. **9**, 102–117 (2015)
8.51 K.W. Boyack, R. Klavans: Including non-source items in a large-scale map of science: What difference does it make?, J. Informetr. **8**, 569–580 (2014)
8.52 R. Klavans, K.W. Boyack: Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge?, J. Assoc. Inf. Sci. Technol. **68**(4), 984–998 (2017)
8.53 J.G. Foster, A. Rzhetsky, J.A. Evans: Tradition and innovation in scientists' research strategies, Am. Soc. Rev. **80**(5), 875–908 (2015)
8.54 F. Shi, J.G. Foster, J.A. Evans: Weaving the fabric of science: Dynamic network models of science's unfolding structure, Soc. Netw. **43**, 73–85 (2015)
8.55 I. Wesley-Smith, C.T. Bergstrom, J.D. West: Static ranking of scholarly papers using article-level Eigenfactor (ALEF). In: *9th ACM Int. Conf. Web Search Data Mining* (ACM, San Francisco 2016)
8.56 J.D. West, I. Wesley-Smith, C.T. Bergstrom: A recommendation system based on hierarchical clustering of an article-level citation network, IEEE Trans. Big Data **2**(2), 113–123 (2016)
8.57 K.W. Boyack, M. Patek, L.H. Ungar, P. Yoon, R. Klavans: Classification of individual articles from all of science by research level, J. Informetr. **18**(1), 1–12 (2014)
8.58 K.W. Boyack, R. Klavans, H. Small, L. Ungar: Characterizing the emergence of two nanotechnology topics using a contemporaneous global micro-model of science, J. Eng. Technol. Manag. **32**, 147–159 (2014)

8.59 L. Waltman, K.W. Boyack, G. Colavizza, N.J. van Eck: A principled approach for comparing relatedness measures for clustering publications. In: *16th Int. Conf. Int. Soc. Scientometr. Informetr., Wuhan* (2017) pp. 691–702

8.60 M.-H. Feng, K.-H. Chan, H.-Y. Chen, M.-F. Tsai, M.-Y. Yeh, S.-D. Lin: An efficient solution to reinforce paper ranking using author/venue/citation information – The winner's solution for WSDM Cup 2016. In: *9th ACM Int. Conf. Web Search Data Mining* (ACM, San Francisco 2016)

8.61 K.W. Boyack, R. Klavans: Measuring science-technology interaction using rare inventor-author names, J. Informetr. **2**, 173–182 (2008)

8.62 S. Reardon: Text-mining offers clues to success, Nature **509**, 410 (2014)

8.63 K.R. McKeown, H. Daume III, S. Chaturvedi, J. Paparrizos, K. Thadani, P. Barrio, O. Biran, S. Bothe, M. Collins, K.R. Fleischmann, L. Gravano, R. Jha, B. King, K. McInerney, T. Moon, A. Neelakantan, D. O'Seaghdha, D. Radev, C. Templeton, S. Teufel: Predicting the impact of scientific concepts using full text features, J. Assoc. Inf. Sci. Technol. **67**(11), 2684–2696 (2016)

8.64 V.I. Torvik, N.R. Smalheiser: Author name disambiguation in MEDLINE, ACM Trans. Knowl. Discov. Data **3**(3), 11–40 (2009)

8.65 G.-C. Li, R. Lai, A. D'Amour, D.M. Doolin, Y. Sun, V.I. Torvik, A.Z. Yu, L. Fleming: Disambiguation and co-authorship networks of the U.S. patent inventor database (1975–2010), Res. Policy **43**, 941–955 (2014)

8.66 W. Liu, R.I. Dogan, S. Kim, D.C. Comeau, W. Kim, L. Yeganova, Z. Lu, W.J. Wilbur: Author name disambiguation for PubMed, J. Assoc. Inf. Sci. Technol. **65**(4), 765–781 (2014)

8.67 C. Schulz, A. Mazloumian, A.M. Petersen, O. Penner, D. Helbing: Exploiting citation networks for large-scale author name disambiguation, EPJ Data Sci. **1**, 11 (2014)

8.68 E. Caron, N.J. van Eck: Large scale author name disambiguation using rule-based scoring and clustering. In: *Proc. Sci. Technol. Indic. Conf., Leiden* (2014) pp. 79–86

8.69 B.D. Fegley, V.I. Torvik: Has large-scale named-entity network analysis been resting on a flawed assumption?, PLoS ONE **8**, e70299 (2009)

8.70 B. Uzzi, S. Mukherjee, M. Stringer, B. Jones: Atypical combinations and scientific impact, Science **342**, 468–472 (2013)

8.71 K.W. Boyack, R. Klavans: Atypical combinations are confounded by disciplinary effects. In: *19th Int. Conf. Sci. Technol. Indicat* (2014)

8.72 V. Larivière, S. Haustein, K. Börner: Long-distance interdisciplinarity leads to higher scientific impact, PLoS ONE **10**, e122565 (2015)

8.73 L. Kay, N. Newman, J. Youtie, A. Porter, I. Rafols: Patent overlay mapping: Visualizing technological distance, J. Assoc. Inf. Sci. Technol. **65**(12), 2432–2443 (2014)

8.74 E.M. Talley, D. Newman, D. Mimno, B.W. Herr, H.M. Wallach, G.A.P.C. Burns, A.G.M. Leenders, A. McCallum: Database of NIH grants using machine-learned categories and graphical clustering, Nat. Methods **8**(6), 443–444 (2011)

8.75 M. Bertin, I. Atanassova, Y. Gingras, V. Lariviere: The invariant distribution of references in scientific articles, J. Assoc. Inf. Sci. Technol. **67**(1), 164–177 (2016)

8.76 K.W. Boyack, N.J. van Eck, G. Colavizza, L. Waltman: Characterizing in-text citations in scientific articles: A large-scale analysis, J. Informetr. **12**(1), 59–73 (2018)

8.77 S. Emmons, S. Kobourov, M. Gallant, K. Börner: Analysis of network clustering algorithms and cluster quality metrics at scale, PLoS ONE **11**(7), e0159161 (2016)

8.78 R. Klavans, K.W. Boyack: Research portfolio analysis and topic prominence, J. Informetr. **11**(4), 1158–1174 (2017)

8.79 D.J. de Solla Price: *Little Science, Big Science* (Columbia Univ. Press, New York 1963)

8.80 K.W. Boyack: Thesaurus-based methods for mapping contents of publication sets, Scientometrics **111**, 1141–1155 (2017)

8.81 R. Klavans, K.W. Boyack: The research focus of nations: Economic vs. altruistic motivations, PLoS ONE **12**, e169383 (2017)

8.82 J.J. Franklin, R. Johnston: Co-citation bibliometric modeling as a tool for S&T policy and R&D management: Issues, applications, and developments. In: *Handbook of Quantitative Studies of Science and Technology*, ed. by A.F.J. van Raan (Elsevier, Amsterdam 1988) pp. 325–389

8.83 R. Klavans, K.W. Boyack: Mapping altruism, J. Informetr. **8**, 431–447 (2014)

Kevin W. Boyack

SciTech Strategies, Inc.
Albuquerque, NM, USA
kboyack@mapofscience.com

Kevin W. Boyack has been President of SciTech Strategies, Inc. since 2007. He received a PhD in Chemical Engineering in 1990. Since joining SciTech his work has centered on developing more accurate global maps of science. He has published over 50 articles dealing with science mapping and related metrics. His current interests include understanding the structure and dynamics of science, technology, and the missions and motives that drive them.

Richard Klavans

SciTech Strategies, Inc.
Wayne, PA, USA
rklavans@mapofscience.com

Dick Klavans is the founder of SciTech Strategies, Inc. He received an Engineering degree from Tufts, a Master's degree from Sloan/MIT, and a PhD from Wharton/U Penn. He has published extensively on the art and science of science mapping (visual representations of the structure and dynamics of research). These maps are primarily for research planning in industry (e.g., Abbott Labs, Astra Zeneca) and for government agencies (e.g., DOE, NSF).

9. Science Mapping and the Identification of Topics: Theoretical and Methodological Considerations

Bart Thijs

This chapter focusses on the drivers for the advancement of mapping of science and the detection of topics as often applied in scientometrics. The chapter identifies three different drivers for this advancement: technological innovation resulting in increased computational power, the improved community detection approaches available today, and advancements in scientometrics itself with respect to the actual linking of documents through citations or lexical approaches. We will show that the main drivers are the first two, with the last one somewhat lagging behind. Next, severe methodological issues have been identified in network science related to the application of these techniques for community detection. The resolution limit and the degeneracy problem are described. The last section shows how different approaches are taken to enable scientometricians to create global maps of science and how they come to comparable results at higher levels of granularity but that the validity of more fine-grained clusters and topics suffers strongly in the discussed problems, which raises serious questions with respect to the applicability of these global techniques with a strong local focus.

9.1 General Drivers for Advancement of Science Mapping

This chapter focusses on the different techniques and methodologies used in bibliometrics to map the cognitive structure of science and the use of clustering techniques for the detection of topics and topical emergence. Traditionally, these mapping exercises are carried out with citation links or textual approaches as the basis for such analysis. But methodologically speaking, it builds strongly on the developments in network science. This chapter will focus on both fields. The first field is of course our own field of bibliometrics, which already has a rich tradition in science mapping, where many studies have been done to suggest new and improved approaches or to compare the different approaches. The focus in these comparative studies is on the type of relations between documents, be it citation-link-based or lexical approaches, but also combinations of both—hybrid networks—have been proposed. Additional work has been done on the labeling of clusters and measurements for the accuracy of the obtained results.

The other field is network science with its roots in both graph theory in mathematics and physics and in social network analysis as part of sociology. The focus of research related to partitioning or community detection is mainly on the development of faster and more accurate algorithms. Finding the best clustering among all

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_9

possible partitions is considered to be NP-hard ([9.1] or [9.2]), which makes it nearly impossible to find an exact partitioning for large-scale graphs. The approach taken by many researchers is to introduce heuristics which exploit some of the properties of the network and that point the algorithm into a particular direction in its search for the optimal solution. Other papers in this field discuss the shortcomings of different approaches or present improvements to existing algorithms to overcome these shortcomings which include the resolution limit [9.3] or the degeneracy problem [9.4]. There is a link between bibliometrics and network science as studies in the latter often use data sets extracted from bibliographic databases.

A third strong driver is the technological advancement in computational resources available to the bibliometrician. In line with Moore's law [9.5], CPU capacity has grown continuously and still is. In addition, the growth in hard disk capacities and the recent introduction of solid state drives combined with the available memory has resulted in a reasonably priced desktop computer capable of the processing of large-scale networks [9.6]. A further advancement has been the development of distributed online cloud computing facilities with enormous computational and storage capacities. Such facilities are publicly available through a renting service, charged by the hour at the fraction of the cost required for the acquisition of such computational power.

Figure 9.1 presents the interplay between these three drivers and their influence on the development of science mapping. The technological development not only enables the analysis of larger document sets but also drives the development of new and improved algorithms. These algorithms are adapted by science mappers but also give a boost to novel research on approaches for the construction of networks amongst scientific papers. The *arrows* in Fig. 9.1 indicate the input of a driver for the advancement of science mapping and its size is relative to its effect. Technological development is the strongest driver, mainly because of the speed of its growth. Next comes the network science input. Newly proposed techniques such as Louvain clustering [9.8], Smart Local Moving [9.2], or the Map Equation [9.7] are rapidly incorporated into the knowledge space of our field.

Much slower, however, is the advancement in scientometrics with respect to the construction of the underlying data sets [9.9]. Citation links were introduced by pioneers like *Garfield* [9.10], *Small* and *Griffith* [9.11] or *Kessler* [9.12]. But after this early emergence of citation links, its development did not assume a fast-growing pace. Studies about the appropriate thresholds can be found along with some applications of second-order similarities. Lexical similarities and hybrid approaches were only added to science mapping in the last decade of the twentieth century. It is only recently that techniques from computational linguistics using natural language processing have been deployed for the creation of document networks.

The slow pace of development in scientometrics as a driver of evolution in science mapping has not prevented this application from quickly adopting to and absorbing the latest advancements in the two other

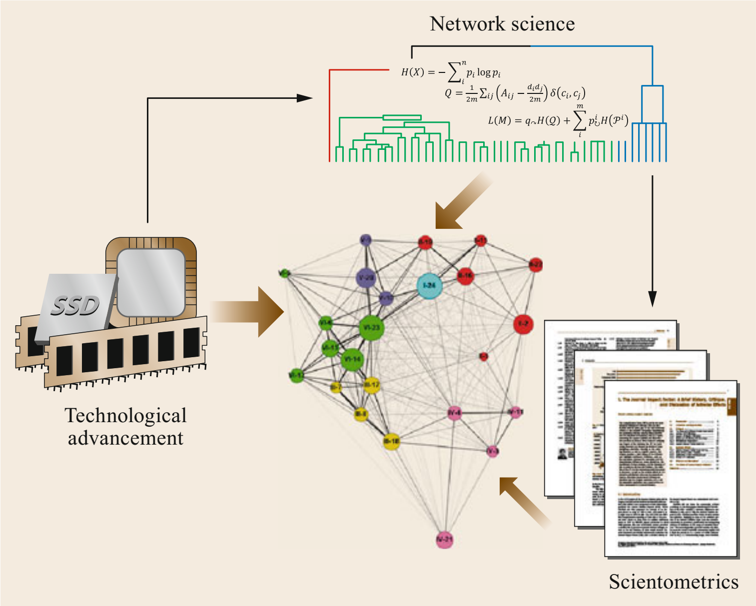


Fig. 9.1 Interaction between and role of the three drivers for advancement in Science Mapping; formulas taken from [9.7]

drivers. The dynamics in the domain of science mapping has been very well described by *Klavans* and *Boyack* [9.13] in their study on the most accurate taxonomy of scientific knowledge.

The objective of this chapter is to provide the reader with a summary overview of the current state of the art of science mapping and topic detection while integrating both the developments in network science and bibliometrics. The main question that will be answered in this chapter is thus:

> How can science mapping benefit from advancements in these two fields while balancing between complexity and accurateness and without losing sight of the requirements set by the different applications?

In other words: how can a scientometrician with a focus on local phenomena whether at individual, institutional or regional level benefit from the latest developments with respect to topic detection?

Another approach to the identification of topics or delineation of subfields is provided by an application that is commonly known as topic modeling which uses hierarchical probabilistic models like latent Dirichlet allocation [9.14], which are based on the assumption that documents can cover more than one but a limited set of topics and that documents use only a limited set of words that are associated to these topics. Such approaches take a term-by-document matrix and result in a term-by-topic model or structure.

In the remainder of the chapter, I will first discuss the development and state of the art of bibliometrics related to the creation of document networks using link-based, lexical and hybrid similarities. Next, community detection techniques commonly used in science mapping are touched upon. After this, methodological constraints related to community detection are discussed together with their implication to the partitioning of document networks. The last section integrates the prior sections in a discussion of local versus global clustering.

The publication data used in this chapter was extracted from the 25 annual volumes of Clarivate Analytics Web of Science (WoS) Core Collection ranging from 1991 up to 2015. Only citable documents are retained and citation links are processed through the custom reference identifiers supplied by Clarivate Analytics. These identifiers can be assigned both to indexed and nonindexed publications.

9.2 Creation of Document Networks

This section starts with a formal definition of a network and then discusses the different approaches taken in the scientometric literature for the construction of document networks.

A graph or network can formally be defined as $G=(N,E)$, being a pair of two distinct sets of information [9.15]. The first set contains all the nodes, whilst the second set holds all the edges. Depending on the topology of the network, edges can be restricted to unordered pairs of elements from set N for undirected unweighted networks, ordered pairs for directed versions or triplets having two elements from N and the third element being a numeric value indicating the strength of the link between the two nodes.

Bibliographic databases allow the creation of networks using different types of nodes. Collaboration studies have been using countries, regions, institutes and individual researchers as nodes. Especially author collaboration networks can hint towards some topical structure in a field. But, as more senior researchers often cannot be pinned to a single topic or even single domain, the reliability of such collaboration networks for topic detection is rather low. At the beginning of this Chapter, alternative methods focussing on aggregation or grouping of the terms are mentioned. Author Cocitation Analysis [9.16] links authors that are cocited to reveal the cognitive structure of a domain. But this chapter will focus on the creation and use of document networks. Attributing documents to clusters allows later a straight forward assignment at higher levels of aggregation. Authors, institutes or journals can then be profiled by the share of their papers among the different clusters. Such a fractionated profile pays tribute to the multitopical multidisciplinary activities of the actor. Hard clustering of entities at these higher levels of aggregation will always hamper the creation of such multifaceted profiles.

9.2.1 Citation-Based Links

Three different types of citation-based links are commonly used in the scientometric literature for science mapping.

Direct Citations

The most straightforward method of linking publications is the use of the actual references found in the citing publication. These references are indeed directed edges from one particular node to another. But the links can be converted into undirected edges which results in an unweighted, undirected network. Such networks have been used by *Ludo Waltman* and *Nees Jan Van Eck* [9.2, 17, 18] who clustered about 10 million papers and by *Kevin Boyack* and *Richard Klavans* [9.13, 19] who were able to produce a map of 43 million documents. Both networks have degrees (number of edges associated with a node) that range around 10. This degree highly depends on the selected publication window and selected publication source. Table 9.1 gives some statistics for publication sets extracted from the Web of Science over different publication windows. The top row shows that 2015 contains 1.6 million documents and less than 1 million links. Taking into account that the link in an undirected network is counted for both nodes for the degree calculation we observe an average degree of 1.13. This degree increases with the extension of the publication window, but the growth has a slow decay. The density of the network given by this formula $2l/(n(n-1))$, with l the number of links and n the number of publications, is extremely low but rather stable, and the least sparse network can be found in a 15-year window.

The low degree indicates the low appropriateness of direct citation (DC) link models for the clustering of short publication windows.

An important conceptual remark has to be made related to the citation behavior of researchers which might have an influence on the applicability of direct citations. Authors referring to prior literature have to be selective as they cannot cite all relevant literature on the studied topic. This means of course that the direct citation link among two documents is created intentionally by the citing authors. The probability of false-positive citation links can thus be discarded. However, the choice of cited literature is subject to affinity to a particular community or even a school of thought which clearly results in false-negatives.

Cocitation

Small and *Griffith* [9.11] introduced a new document analysis approach based on cocitation which was defined as the frequency at which two documents are cited together. They used this model for the creation of network models with the objective of studying the structure of science. The patterns they observed using cocitations were highly similar to those obtained using direct citations and differed significantly from the bibliographic coupling (BC) patterns. Small and Griffith decided to use cocitation because that allowed them to restrict their analysis to publications within a single year without the requirement to include all cited documents [9.13].

Cocitation analysis has been regularly used for the detection of research fronts due to the dynamic component that is inherently connected to this model. The relation between documents can change over time if the source time window is shifted.

Boyack and *Klavans* [9.19] put forward four main features of cocitations in the modeling of science which they collected from prior literature:

1. They could show that the clusters generated using cocitations are sufficiently accurate [9.20].
2. Because of the weighted nature of cocitation links, papers can be fractionally assigned to clusters. And clusters can be linked through paper overlap [9.21].
3. Cocitation analysis tends to produce unstable solutions which is regarded as being very useful for the detection of emerging topics [9.22].
4. The study of the dynamics of science benefits from the longitudinal structures that characterize the cocitation clusters [9.23].

Bibliographic Coupling

Two papers are bibliographically coupled when they share at least one reference, see for example [9.12, 24]. When we consider references as an indication of use of a particular set or item of knowledge we can infer that papers that share a reference also have some common use of prior knowledge and share a partial topical similarity. The problem of the possible false-negatives in citation links also affects the bibliographic coupling links between documents, especially if community- or

Table 9.1 Number of publications and links, both expressed in millions (m) and average degree and density networks over different publication windows

Publication window	Number of publications (in millions)	Number of links (in millions)	Average degree	Density
2015–2015	1.6	0.9	1.13	7.08×10^{-7}
2013–2015	4.6	13.1	5.65	1.22×10^{-6}
2011–2015	7.4	38.9	1.47	1.41×10^{-6}
2006–2015	13.3	140	20.99	1.57×10^{-6}
2001–2015	18.0	256	28.44	1.58×10^{-6}
1996–2015	22.2	362	32.54	1.46×10^{-6}
1991–2015	26.0	448	34.50	1.33×10^{-6}

Data sourced from Clarivate Analytics Web of Science Core Collection

field-specific motivations drive the selection of cited papers.

Comparisons between cocitation and bibliographic coupling can be traced back to as early as the *Small* and *Griffith* paper [9.11], which proved the distinct nature of this latter approach as opposed to the other citation-based links. In 1996, Glänzel and Czerwon pointed out that almost all relevant papers in a document set have references and that these references are present at the time of publication which means that bibliographic coupling links can be created as soon as a paper is published. Cocitation links only appear after a paper gets its first citation.

The large amount of data that can be used as the source for the creation of bibliographic coupling networks is illustrated by Table 9.2. For an increasing publication window from one up to ten years, a set of statistics are provided. In the full ten-year window from 2006 up to 2015, more than 60 million different cited documents can be identified in the Web of Science (not all of them are indexed by the WoS). About 48% of these documents are mentioned in more than one indexed publication and can thus be used to link these publications. On average, each reference is cited 12.8 times in the ten-year window. But as citation distributions are very skewed, the most frequently cited reference has received 45 147 citations in the database that was used. This is the paper by *Sheldrick* [9.25] about a software package called *SHELX*, which is commonly used in crystallography.

Dealing with such a high number of citing papers requires a performant processing environment. In the past, several researchers have surpassed this problem by the application of certain thresholds. References were removed from the set if they were cited more than a fixed number of times. Recently, I presented a more sophisticated methodology which allows the choice of selection criteria and threshold at the level of the reference itself. This proved to improve the scalability of the creation of bibliographic coupling networks and showed that removing the most cited references from the set results in the absence of many small BC-links between documents while the removal of the large set of references that are only cited a few times hampers the detection of strong links between documents [9.26].

9.2.2 Lexical Similarities

Textual analysis became part of the bibliometricians science mapping toolbox much later. It was *Callon* et al. [9.27] who was the first to demonstrate the usability of coword analysis. An important reason for this lagging behind the link-based approach is the late digital availability of large text corpora data and the extensive computational requirements. In Leiden, *Tijssen* and *Van Raan* [9.28] used co-occurrence of keywords in patents and publications for the detection of links between science and technology. In Leuven, *Glenisson* et al. [9.29, 30] started experimenting with full text retrieved from scientific papers instead of terms extracted from keywords, titles or abstracts. They concluded that the preference for full text over title and abstract approaches is not as yet clear and still open to discussion.

The creation of a document network using lexical similarities requires a much more elaborated approach than the citation link counterpart. There are no direct, author-chosen textual links between documents. This implies that any lexical link or similarity between two documents has to be deduced from the actual content of both documents. The science mapper has to use different techniques to extract textual information, match documents with each other and calculate lexical similarities. Many of these techniques were developed in the field of information retrieval and in (computational)

Table 9.2 Descriptive statistics of references used as the source for bibliographic coupling counts expressed in millions

Publication window	Unique reference count (in millions)	Linking references (in millions)	Share of linking references (%)	Average number of documents linked	Occurrence of most-frequent references
2006	11.5	4.69	40.8	3.80	3826
2006–2007	17.5	7.95	45.4	5.33	7711
2006–2008	20.3	10.9	47.0	6.68	9598
2006–2009	28.5	13.6	47.8	7.82	12 424
2006–2010	33.4	16.0	48.0	8.81	15 932
2006–2011	38.5	18.5	48.0	9.72	23 668
2006–2012	43.9	21.1	48.0	10.58	29 159
2006–2013	49.2	23.6	48.0	11.35	34 172
2006–2014	54.9	26.3	48.0	12.12	39 382
2006–2015	60.7	29.1	48.0	12.80	45 147

Data sourced from Clarivate Analytics Web of Science Core Collection

linguistics and have been adapted for application in the specific context generated by scientific literature. Before a text-based document network can be built the process has to pass through four major steps:

1. Source selection
2. Preprocessing
3. Mathematical representation
4. Document linking.

Choices made during each of these steps have consequences for the final result of the clustering of documents and scientometricians should be well aware of the impact of their choices on the complexity of the analysis and the accuracy of the identified topics in the document sets.

Source Selection

The augmented complexity of the lexical similarity already starts with the selection of the source data to be used for the creation of the document network. Of course, first and completely in line with the citation-link approach, the relevant document set has to be delineated. But after that, an abundance of possible choices is present to the science mapper. The lexical similarity can be based on the title [9.10] keywords (given by the author or editor) [9.31], on extended versions of keywords like Keyword Plus provided by Web of Science, on specialized headings [9.32] or terminology like the Medical Subject Headings (MeSH). More comprehensive are lexical similarities based on the combination of title and abstract or on full texts of papers [9.29]. The problem with the last option is the lack of availability of large sets of digital full text documents covering a complete topic. Often, only a limited selection of papers within a domain are freely available.

Both Scopus and Web of Science provide the title and abstract of nearly all papers indexed together with the different versions of keywords. Several authors have published a small, local mapping exercise using full text for which they downloaded all relevant full papers from the journals websites or from online repositories. Nowadays, authors tend to make use of the possibility offered by several publishers to self-archive draft or prefinal versions of their own papers. These papers can then be automatically retrieved and used to enrich the bibliographic database. Dealing with the extraction of relevant text from files in different formats like PDF, XML or HTML is not straight forward. The availability of several (well-priced) commercial products proves the complexity of automating this task.

Preprocessing

After the source is selected and the data is retrieved, the preprocessing phase can start. The different tasks in this step are common to most text processing projects. Text indexing software packages like Lucene (https://lucene.apache.org/core/) include tools for tasks such as segmentation and tokenization, filtering, stemming and analyzing.

The preprocessing starts with the extraction of relevant terms from the digital document files or database records. In the case of keywords and subject headings like MeSH terms, only a little further preprocessing is required, i.e., conversion to lower case, removal of specific tags or codes [9.33].

For title and abstracts, a more complex procedure should be followed. The text must be tokenized using white space and punctuation. Numbers or words containing digits may be removed.

Stop words are words that are very common and bear no specific meaning or semantic value and should be removed during preprocessing. Many software packages provide their default stop word lists which can be extended when needed.

Stemming is a process that reduces the term to its root form. Because of inflection, which expresses differences in e. g., number, gender, person or tense, terms appear in different forms across texts. Lexical similarity between documents will be underestimated if no stemming is applied. The best-known stemmer is the Porter stemmer [9.34], using a simple rule-based scheme.

Both stemming and stop word removal have the additional advantage that they reduce the dimensionality of the vector space.

After all this preprocessing, *Glenisson* et al. [9.29] still encountered some disadvantages for single-term approaches and he added a set of 500 bigrams selected by using the Dunning Likelihood ratio test [9.35]. These bigrams were among the most frequent occurring combinations of two terms. However, this bigram-based solution only increased the computational complexity without clear added benefits.

Leopold et al. [9.36] mentioned the possible advantages of the application of Natural Language Processing (NLP), which incorporates structural properties of the text over a bag-of-words approach where all words are treated equally and independently from each other. At that time, 2004, they thought it was impossible to implement such an NLP-framework for a large amount of texts with the techniques and computational resources available then ([9.36, p. 202]) and they suggest Shallow Parsing as a more pragmatic alternative [9.37]. One decade later, *Thijs* et al. [9.38] started using Natural Language Processing for the extraction of noun phrases (NPs) from title and abstracts. The main rationale behind this approach is that when using text mining for science mapping the textual information should be restricted to those parts that actually reflect topics and concepts. Syntactic parsing enables the extraction of noun phrases while neglecting verbs and adpositional phrases. In this paper Thijs et al. [9.38] used the Stanford Core NLP package. Figure 9.2 gives the output from parsing the following sentence obtained from a sample publication in the document set:

> Results of the study show that information systems downsizing may produce benefits such as improved information systems, improved organizational structure, higher productivity, and lower cost. [9.39]

Only those elements that are tagged by NP are retained for the analysis and after tokenization, stemming and removal of stop words, term shingles were created which contain at least two distinct terms.

Retaining all noun phrases that contain only a single term would be nothing more than a further extension of the bigram approach taken by *Glenisson* et al. [9.29] and would not significantly contribute to the improved accuracy and applicability of the lexical component in document clustering.

The NLP approach also proved very successful while analyzing large document sets. The technique was used for the detection of topics in the astronomy dataset holding more than 110 K documents within reasonable time and space constraints [9.40].

Mathematical Representation

The most common approach to create a mathematical representation of a document set is to adopt a vector space model to encode each document using a k-dimensional vector with k equal to the total number of terms or phrases present in the total set of documents. Such a model results in a sparse document-by-term matrix as most documents will only contain a limited set of the terms present in the complete set. The cells in the matrix can contain binary values indicating the presence/absence of a particular term in the document but the value can also be an integer indicating the number of occurrences of the term in the document or even a real value with a weighted or normalized term frequency.

The standard Term Frequency-inverse Document Frequency (TFiDF) as proposed by *Salton* and *Buckley* [9.41] is given by the formula

$$w_{t,d} = \mathrm{tf} - \mathrm{idf}_{t,d} = \mathrm{tf}_{t,d} \times \mathrm{idf}_t \,,$$

which gives a weight to term t in document d. High TFiDF values are obtained for terms occurring several times in a document but only a limited number of times across different documents.

TFiDF is a typical bag-of-words approach as each term is processed independently from the other terms in the document or independently from its morphological function in the sentence. Ordering of the words in the

Label	Result of the Stanford Parser
	`(ROOT`
	`  (S`
A	`    (NP`
A1	`      (NP (NNS Results))`
	`      (PP (IN of)`
A2	`        (NP (DT the) (NN study))))`
	`    (VP (VBP show)`
	`      (SBAR (IN that)`
	`        (S`
B	`          (NP (NN information) (NNS systems) (NN downsizing))`
	`          (VP (MD may)`
	`            (VP (VB produce)`
C	`              (NP`
C1	`                (NP (NNS benefits))`
	`                (PP (JJ such) (IN as)`
C2	`                  (NP`
C2a	`                    (NP (VBN improved) (NN information) (NNS systems))`
	`                    (, ,)`
C2b	`                    (NP (VBN improved) (JJ organizational) (NN structure))`
	`                    (, ,)`
C2c	`                    (NP (JJR higher) (NN productivity))`
	`                    (, ,)`
	`                    (CC and)`
C2d	`                    (NP (JJR lower) (NN cost)))))))))))`

Fig. 9.2 Syntactic tree as a result of parsing the example sentence

sentence or in the document is completely lost. The approach for the extraction of noun phrases from titles and abstracts developed by *Thijs* et al. [9.42] moves away from this bag-of-words framework as terms and phrases are selected based on their function in the sentence and the order of the terms is retained to some extent through the use of shingles.

Special care has to be taken for those terms or phrases that occur in only one document. They can be excluded from the vector space. Table 9.3 gives some statistics about the number of noun phrases in the Web of Science publication set from 2006 up to 2015. About 13% of the noun phrases extracted from the publications using the above-described methodology are shared among at least two documents. This share is much lower than the 48% found in Table 9.2 for cited references. The average document frequency for the shared NPs ranges from 7.45 up to 12.02. These are still reasonable values. However, when looking at the most common noun phrase (`long term`) we see very high values. This noun phrase occurred in more than 340 000 documents.

Other very common (stemmed) noun phrases include: `follow up`, `signific differ`, `electron microscopi`, `signific higher`, `significant increase`, `important role`, `cross section`, `risk factor`, `ray diffract`, `year old`, `statist signific`, `first time`, `confid interval`, `real time`, `gene express`, `scan electron`, `three dimension`, `control group`, `cell line`, `present paper`.

The presence of such common words is completely in line with Zipf's law and indicates the importance of the weighting factor, the inverse document frequency [9.43]. There are several approaches for the calculation of the iDF. An implementation grounded in information theory is given by this formula

$$\log\left(\frac{N}{\|\{d \in D : t \in d\}\|}\right)$$

which is the log of the inverse of the probability that a randomly picked document d contains the term t and thus related to Shannon's entropy or the average information contained in the fact that term t can be found in document d. The compliance of the frequency distribution to Zipf's law results in the sparse data problem [9.36]. Documents, and especially the combination of title and abstract, contain only a very limited set of terms identified in the complete data set. Consequently, the vectors containing the mathematical representation of the document will contain mainly empty or null-valued cells.

A common solution applied in data mining is to reduce the hyperdimensionality of the vector space model with nearly 70 million different noun phrases through the application of dimensionality reduction techniques like PCA (principal component analysis), which is based on the covariance matrix, or SVD (singular value decomposition) and LSA (latent semantic analysis) which take the source matrix. These techniques do not take the skewed frequency distribution into account and might result in less accurate results. A method that is more suited to text mining is probabilistic latent semantic analysis [9.44]. Other techniques like Random Project, which are less computationally demanding, are applied to larger data sets [9.45].

Document Linking

The last step in the creation of the document network is linking documents to each other.

Salton proposed a *cosine similarity* score for document retrieval defined as the dot product between the two vectors in the VSM (vector space model) associated with a document and a particular query. This can easily be adapted to a score between two documents in a document space

$$\operatorname{sim}\left(d_i, d_j\right) = \frac{\sum_{t=1}^{N} w_{t,i} w_{t,j}}{\sqrt{\sum_{t=1}^{N} w_{t,i}^2}\sqrt{\sum_{t=1}^{N} w_{t,j}^2}} .$$

Another function taken from information retrieval is the *Okapi BM25* ranking function (or short BM25, where BM stands for Best Match) as proposed *Spärck-Jones* et al. [9.46] and applied in science mapping by *Boyack* et al. [9.33]. This is a probabilistic model that is sensitive to the size of a document and thus suited for larger documents while keeping the complexity rather low [9.47]. The formula for the calculation of the BM25 similarity is given by

$$\begin{aligned}&\operatorname{sim}\left(d_i, q\right)\\&\quad = \sum_{t=1}^{n} \mathrm{IDF}(q_t)\frac{f_{t,i}\left(k_1+1\right)}{f_{t,i}+k_1\left(1-b+b\frac{\|d_i\|N}{\sum_{j=1}^{N} d_j}\right)}\end{aligned}$$

with typical vales for k_1 and b set to 2.0 and 0.75.

Although the calculation of both cosine as BM25 are straightforward for any given pair of vectors, the task becomes very time consuming with a growing network. In fact, the calculation requires the matching of all possible document pairs available in the network. This number grows quadratically as the number of possible links in a network with n nodes is equal to

$$\frac{n(n-1)}{2}$$

Table 9.3 Descriptive statistics of noun phrases (NP) used as the source for lexical similarities

Publication window	NP count (in millions)	Linking NPs (in millions)	Share of linking NPs (%)	Average number of documents linked	Occurrence of most frequent NP
2006	49.4	6.23	12.6	7.45	25 287
2006–2007	92.1	12.0	13.1	8.53	50 999
2006–2008	140	18.6	13.3	9.29	83 409
2006–2009	188	25.2	13.4	9.86	117 125
2006–2010	236	31.7	13.5	10.31	149 911
2006–2011	286	38.6	13.5	10.71	184 717
2006–2012	340	46.0	13.5	11.09	222 159
2006–2013	395	53.5	13.5	11.42	260 273
2006–2014	456	61.6	13.5	11.73	300 048
2006–2015	519	69.8	13.5	12.02	340 607

Data sourced from Clarivate Analytics Web of Science Core Collection

Tests with a small publication set about 'Information Systems Research' in the social sciences revealed a network with extremely high density of 96% [9.42]. Almost all of the possible links in this network of 6144 documents were present. Similar high density scores are obtained with other text-based networks as most document pairs share at least one or a few very common terms. The application of noun phrase detection as described above solved this issue with the density of the networks dropping drastically.

Next, storing such a lexical network is challenging. In the above example with publication data from 2006 up to 2015, the noun phrase `long term` appeared in 360 607 documents. This term alone creates 6.5×10^{10} document pairs. Assuming that each document is identified by a code of 4 bytes and that the similarity between the documents requires an additional 4 bytes, storing the network of documents connected by the noun phrase `long term` alone would require 78 gigabytes of disk space. The use of a weighting or normalization of the term frequencies by an inverse document frequency does not lower the required storage as only the weight decreases but the link does not disappear. Applying a certain threshold for the weight of the document link might help as it will allow us to remove a large portion of very small lexical links.

However, the proposed methodology would still require a long computation time as the links between documents must first be calculated before they are rejected due to being below a threshold. Recent approaches including the use of Locality Sensitive Hashing [9.48] for lexical approaches or Drakkar for bibliographic coupling [9.26] solve this problem as they use randomization and hashing functions to reduce the number of pairwise comparisons. The reduction is relative to the probability that two documents share terms or references. The higher the probability that two documents share terms or references, the higher the probability that they are assigned the same value by the hashing function in LSH (locality-sensitive hashing).

The complexity involved in the creation of document networks using lexical similarities is indeed a serious burden for the applicability of this approach. *Van Eck* and *Waltman* [9.18] deservedly state that homonyms (similar words with different meaning across topics or fields) or very common or general terms (like `long term`) can distort the analysis.

9.2.3 Hybrid Approaches

It was *Bichteler* and *Eaton* [9.49] who were the first to suggest the combined use of different document similarities (bibliographic coupling and cocitation), although it was in the context of document retrieval. About a decade later, *Braam* et al. [9.50, 51] and *Zitt* and *Bassecoulard* [9.31] proposed the combination of text- and link-based similarities for science mapping. They added coword analysis to improve the efficiency of the cocitation approach.

One of the reasons for the improved performance of the combined approach is that link-based networks tend to be extremely sparse and underestimate the actual links among documents.

Janssens et al. [9.52] combined the distances based on bibliometric features with a lexical weighted link using this formula

$$D^{\mathrm{INTEGR}} = \lambda D^{\mathrm{TEXT}} + (1-\lambda) D^{\mathrm{BIBL}}$$

But the different distributional characteristics of both underlying similarities posed an immediate threat to the applicability of this procedure. *Thijs* et al. [9.38] plotted both distributions for a selected set of papers from information systems research. Figure 9.3 is taken from this paper and shows a clear distinction between the distribution of the weighted degree

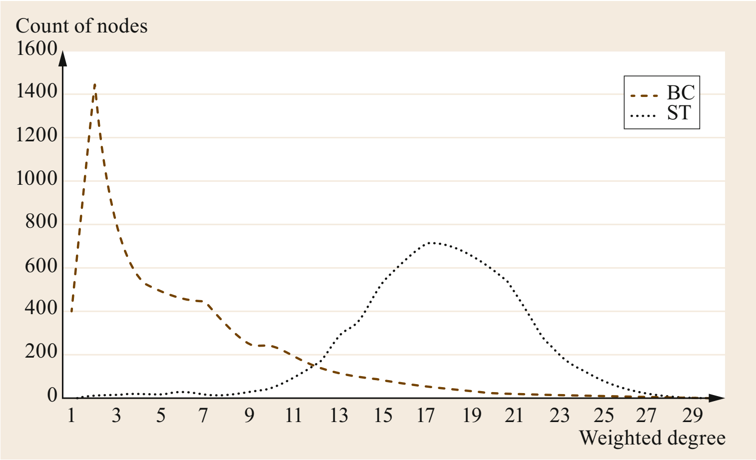


Fig. 9.3 Distribution of weighted degree of single term (ST) and bibliographic coupling (BC) networks

in a bibliographic coupling network and one based on lexical similarities using single terms (ST) for the calculation of Salton's cosine. The latter approximates a normal distribution while the first displays much more scale-free properties commonly found in real world networks [9.53].

A probabilistic method was later proposed by *Janssens* et al. [9.54] where they re-scaled each similarity into the corresponding p-value based on a cumulative distribution function of the similarities in a completely randomized dataset. This method solved the issue of different distributions drastically but it introduced a calculation scheme that was hardly applicable for large networks.

In 2012, we proposed another approach [9.55] where we used the fact that both measures are cosines expressing the similarity or angle between two vectors in a vector space model. In particular, the hybrid similarity measure r is defined as the cosine of the linear combination of the underlying angles between the vectors representing the corresponding documents in the vector space model, i. e.,

$$r = \cos\left(\lambda \cdot \arccos(\eta) + (1-\lambda) \cdot \arccos(\xi)\right) ,\quad \lambda \in [0, 1] ,$$

where η is the similarity defined on bibliographic coupling and ξ the textual similarity, $\arccos(\eta)$ and $\arccos(\xi)$, respectively, denote the two underlying angles. The λ parameter defines the convex combination which allows us to put more weight on one or the other component.

In the first applications of this approach, the λ parameter ranged between 0.75 and 0.875 and was used to improve the clustering results obtained from the bibliographic coupling. It seemed that the hybrid approach with a small portion attributed to the lexical, bag-of-words component was capable of generating more focussed maps than relying solely on the bibliographic coupling. Extensive testing showed, however, that shifting the balance more towards the lexical component would blur the map completely [9.56]. The improvement of the lexical component through the application of NLP changed the interplay between both components drastically. The distributions of the weighted degree of both similarities is now more in line with each other and the risk of distorting the partitioning through the choice of an inappropriate parameter is drastically reduced. Changing the λ parameter is now more analogous to changing the viewpoint on a document set while keeping the focus of your map [9.42].

9.3 Techniques for Community Detection

9.3.1 Linkage and Goodness of Clustering

The absence of a clear definition of communities or clusters supported by scholars across disciplines implies the inability to come up with a set of universal principles to judge the quality of the partitioning. The validity of a given definition for clusters, communities or partitions not only depends on the formal correctness but also on the relevance of the concepts for the actual application in which the community detection is applied. In most general terms one could state that:

> A clustering of nodes in a document network for the identification of topics is successful if papers that deal with the same topic are grouped together and those papers which don't deal with a particular topic are excluded from that cluster.

The success of an approach not only depends on its ability to group papers but also on its discriminative power [9.57]. The solutions should be fine grained enough without generating trivial high-level general clusters like *life sciences* versus *natural sciences*. Such a general statement when formulated more formally by researchers is then used as the basis for the development and implementation of clustering algorithms. *Newman* and *Girvan* [9.1] focussed on the number of links inside a cluster in contrast to the links between clusters. A good clustering partitions the network in such a way that most edges are found inside and not between clusters. To measure the difference between both they introduced the concept of *Modularity*. A completely different approach was proposed by *Rosvall* and *Bergstrom* [9.7] who used concepts from information theory and combined this with the notion of random walks in networks. They introduced the *Map Equation* function which assumes that the optimal clustering of a network allows the shortest coding of any path a random walker takes across the network.

Having a formal definition of a good clustering with an associated formula and implementation does not provide us yet with an algorithm that enables us to obtain this best clustering. Having to calculate the Modularity or Map Equation function for all possible partitions of a network would be nearly impossible, especially for larger graphs. *Newman* and *Girvan* considered this problem to be NP-hard (see also [9.2]). Many different approaches have been developed using heuristics that guide the algorithm towards the optimal solution. This is related to the linkage criteria in more traditional clustering. Most research has been devoted to finding the best heuristics to solve this problem. In what follows, I will provide a more detailed description of both the Modularity and the Map Equation function and then focus on the most common techniques developed for their maximization.

9.3.2 Modularity

Newman and *Girvan* [9.1] defined their measure for the quality of the partitioning of a network (called Modularity) as the difference between the fraction of edges within the clusters and the expected fraction given a complete random distribution of edges in the network. The idea for looking at the difference between strength of the links in subgroups as opposed to those between subgroups dates to *Bock* and *Husain* [9.58] who proposed looking at the ratio of the number of links inside versus the number of links between clusters both normalized by the number of possible links based on the size of the network and the cluster [9.15].

Instead of using a ratio, Newman and Girvan proposed subtracting the expected share of links inside a cluster from the actual observed share. The expected share of links is based both on the links within but also on the links to other clusters. The original formal definition in their 2004 paper is

$$Q = \sum_i \left(e_{ii} - a_i^2\right)$$

Here e_{ii} refers to the fraction of edges that interconnect nodes within a community over the total set of all edges in the network and a_i^2 expresses the expected fraction of inside links. a_i is based on the row-total for all vertices in a cluster in the adjacency matrix. The formula can be rewritten to incorporate this adjacency matrix A.

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{d_i d_j}{2m}\right) \delta\left(c_i, c_j\right) \qquad (9.1)$$

Here m is the total number of edges in the network, d_i is the number of edges for node i (its degree) and $\delta\left(c_i, c_j\right)$ is the *Kronecker delta* being 1 if i and j belong to the same cluster.

Theoretically, the modularity score ranges from -1 up to $+1$ but these are asymptotical values. However, these values are not reached in real world situations and even not in small, perfectly partitioned networks as is demonstrated in Table 9.4. Suppose a network with 5 perfect modules, having only links inside the cluster and none from one cluster to another. The highest possible modularity score only reaches 0.768, far below the theoretical maximum of 1. This value of 1 is only reached with an infinite number of clusters where the expected share is infinitesimally small compared to the observed share.

Table 9.4 Demonstration of the maximum score in a perfectly partitioned network

Cluster	1	2	3	4	5
Number of links	75	100	45	35	45
e_{ii}	0.25	0.33	0.15	0.116	0.15
a_i^2	0.06	0.11	0.02	0.013	0.02
$\left(e_{ii} - a_i^2\right)$	0.19	0.22	0.13	0.103	0.13
Modularity score					0.768

One of the consequences of this practical limit to the maximum modularity score is the incomparableness of scores among different solutions with a different number of clusters.

Newman and *Girvan* [9.1] and later *Newman* [9.57] proposed a divisive approach starting from one large cluster and iteratively splitting clusters up to the point a certain stopping criterion is met. In the first paper, cuts or splits are executed after the calculation of edge betweenness using random walkers. In the second paper, Newman reformulates the modularity function in terms of the spectral properties which reduces the maximization problem to an eigenvalue problem.

A commonly used heuristic for modularity optimization is that of the local mover. The idea is that nodes can only be merged with other nodes or clusters along the edges that connect them. Nodes are thus potentially moved following their edges and after each possible move the change in modularity is calculated. The move with the highest increase in modularity is retained. This procedure is successfully implemented by *Blondel* et al. [9.8] in their *Louvain* method.

Blondel and colleagues proposed a bottom-up linkage procedure where each node starts in its own cluster. Next, changes in modularity are calculated if nodes are merged into neighboring clusters by moving them along the path that connects the node with its direct neighbors. The move that results in the highest increase will be retained. This process is repeated until no further increase can be obtained. The process will now aggregate all nodes within one cluster and creates a new network based on the aggregated nodes or clusters and the intercluster links. The procedure then starts again and moves clusters along their links to neighboring clusters. It is important to realize that the nodes are moved in a sequential order where the sequence in which nodes are considered has an undeniable effect on the result of the clustering. Often, this is dealt with by the implementation of multiple runs of the same algorithm but with a different, randomly organized sequencing of the nodes. Several researchers have since built on the work of Blondel et al. to improve their approach. *Rotta* and *Noack* [9.59] introduced a more effective approach which uses multilevel refinement and *Waltman* and *Van Eck* [9.2] extended the approach by an additional application of the local moving heuristic inside each community during the procedure.

9.3.3 Map Equation

Rosvall and *Bergstrom* [9.7] take a completely different approach when defining criteria for a good clustering. Their key idea is rooted in the seminal work of *Shannon* and *Weaver* [9.60] in information theory which states that a stream of information can be coded in such a way that it capitalizes on the structure or regularities of the underlying system. Rosvall and Bergstrom propose the use of the concept of a random walker across the network for the description of the information flow. The objective is then to find the most optimal coding scheme to describe the path through a network where optimal is considered as being as short as possible. The map equation framework further builds on these assumptions:

- Nodes that are more frequently visited by the random walker get a shorter code
- Codes can be reused among different partitions or modules in the network
- Modules themselves are also coded
- A specific code indicates the exit of a module by the random walker
- In order to avoid death locks of the random walker, a small teleportation probability similar to the one applied in Google's PageRank [9.61] is added.

The direct advantage of the use of a random walker for the description of information flow is that it intrinsically allows the use of both edge directions and edge weights.

In fact, the objective is not to implement a real random walker and find the actual coding of its path but to calculate the expected code length given the underlying network structure, the proposed partitioning and the above-mentioned constraints. Here, Shannon's entropy for a random variable can be used as a lower bound for the average code length.

$$H(X) = -\sum_{i}^{n} p_i \log p_i$$

with X being a discrete random variable having possible values in $\{x_1, x_2, \ldots, x_n\}$ and associated with a probability mass function $P(X)$.

Finally, *Rosvall* and *Bergstrom* [9.7] use a weighted combination of the entropies for the flow of information across the modules and the flow inside each of the distinct partitions.

$$L(M) = q_{\curvearrowright} H(\mathfrak{Q}) + \sum_{i}^{m} p_{\circlearrowright}^{i} H\left(\mathfrak{P}^{i}\right)$$

This formula describes thus the lower bound for the expected code length for a random walker in a network with a partitioning M of m different modules.

As mentioned above, a clustering or community detection algorithm must also implement a linkage scheme. For this, *Rosvall* and *Bergstrom* [9.7] find their inspiration in the local moving heuristics used already by *Blondel* and others [9.8]. For a detailed description of the map equation procedure and application of the framework I refer the reader to the tutorial by *Bohlin* et al. [9.62].

The random walker approach taken in the map equation framework has opened the door for many very interesting but quite natural extensions of this community detection approach:

- Dynamics: Changes in network structures are detected using a two-stage approach. First, clustering is performed for each time slice or publication window and next an alluvial diagram is used to highlight the structural changes [9.63].
- Bipartite networks. Often, clustering approaches to bipartite networks will first try to project the bipartite structure of the network into a unipartite counterpart. Because the map equation is built on a random walker being a Markov process at discrete time, a continuous time generalization was possible [9.64] and an implementation for bipartite networks was developed by *Kheirkhahzadeh* et al. [9.65] without the need for a unipartite projection.
- Overlapping modular organization: Sometimes, nodes are located at the boundaries of clusters or communities and the minimum average code length calculated in the map equation framework benefits from assigning nodes to overlapping clusters where random walkers passing through this node and remaining in the same cluster in the next step never have to be coded as leaving the cluster. Such nodes get distinct codes in each of the overlapping clusters [9.66].
- Multilayered networks. These networks have multiple types of edges or links between the nodes. In the case of the previously discussed hybrid networks there would no longer be the need to merge bibliographic coupling and lexical networks as both links can exist next to each other in a multilayered network. *De Domenico* et al. [9.67] showed the capabilities of the map equation framework on a scientist network using their joint affiliation and their collaboration.

9.4 Methodological Constraints

9.4.1 Resolution Limit

Fortunato and *Barthélemy* [9.3] identified and described an important drawback of clustering based on the optimization of the modularity function. In fact, the approach holds a particular (*intrinsic* as they call it) scale property which depends on the total number of edges or links present in the network. Because of this limit, small clusters with a size close to or below this scale will be neglected in the most optimal solution and be absorbed by other, larger clusters. In extreme cases, the existence of a single link between a small clique and a larger cluster can be sufficient for the phenomena to occur. *Fortunato* and *Barthélemy* [9.3] coined the term resolution limit to denote this problem and they pointed out that it arises from the fact that the optimization is based on a sum of terms where one has to balance between many lower terms (many modules with lower modularity) and a few terms with higher values. They concluded that modularity-based methods tend to have some a priori set, mathematically defined optimal solution which is possibly not in line with the real structure of the underlying network. The problem is most likely to happen in large networks with diverse cluster sizes.

Finally, Fortunato and Barthélemy make a strong point against the use of quality functions that are based on a global model for setting the expected share of links and they advocate for a local definition of the quality of a good partitioning.

Blondel et al. [9.8] argue that the Louvain community detection approach does not suffer from the resolution limit due to the multilevel nature and due to the local moving heuristic; nodes are moved around but it is very unlikely that all nodes from one cluster are moved towards a neighboring cluster. In a later stage, clusters might indeed be merged together, maybe due to the resolution limit. *Van Eck* and *Waltman* [9.18] use the argumentation from *Traag* et al. [9.68] to claim that their approach does not suffer from the resolution limit as they have replaced the expected share of links from the global random graph model by a general resolution parameter.

Kawamoto and *Rosvall* [9.69] estimate the resolution limit of the map equation framework and show that the problem is much smaller than for modularity-based approaches. However, when the network becomes large enough the detected community structure will suffer from the resolution limit. This is in line with the findings from *Traag* et al. [9.68] who refer to the use of global quality functions as a general cause.

Related to the resolution problem, *Lancichinetti* and *Fortunato* [9.70] discussed another, strongly related, problem. Resolution limit describes the problem of absorbing smaller clusters into larger ones. But at the same time, the opposite problem might occur: modularity maximization approaches split clusters. Lancichinetti and Fortunato could prove that it is impossible to avoid both problems at the same time and that it is related to the optimization of a global quality function analogous to the conclusions that *Traag* et al. [9.68] reached at nearly the same moment. Even a multilevel approach or the introduction of a resolution parameter does not resolve the issue and is insufficient to make the community detection approach insensitive to one or both problems.

9.4.2 The Degeneracy Problem

Beside the widely studied problem of the resolution limit, *Good* et al. [9.4] describe two additional issues which modularity maximization suffers from. These are related to the general acceptance of several assumptions underlying the practical application of such methods.

Good and his colleagues [9.4, p. 2] listed three of them:

1. It is possible to find the optimal partitioning for real networks with a clear community structure.
2. Other partitioning's with high scores on the modularity function will have similar community structure as the optimal clustering.
3. The obtained modularity score can be compared across networks.

But, actually, none of these assumptions hold in real world networks. First, they could prove that a large number of possible alternative solutions have modularity scores that approach the highest value (the optimum) while providing clearly deviating partitions. A plot of modularity scores is not topped by a single peak but rather an irregular shaped plateau, often called *glassy* in physics literature. Any algorithm could easily get caught in one of the many small spikes that constitute this typical form. Moving away from this spike will result in a decrease of the total modularity score and brings the algorithm to a halt. This issue is called the *degeneracy problem* in community detection and places a serious burden on the validity of any single partitioning result. This same problem is also recognized by *Lancichinetti* and *Fortunato* [9.70] where they point out that it can be quite easy to find a partition with a modularity score that is quite close to the optimal value but that at the same time it is impossible to find this optimal solution.

Next, they noted that the maximized modularity score for a particular network cannot be compared with other scores as the obtained value does not solely depend on the quality of the partitions but also on the size of the network and on the number of modules.

They close their paper with a strong caution warning users that community detection using global optimization techniques are sensitive to problems that are caused by incorrect merging and splitting of clusters. They see these global techniques only performing well with networks consisting of clusters of similar size while this is especially not the case in large-scale networks [9.70].

A common approach to overcome these issues is the application of multiple runs of the same algorithm with different initialization parameters. Better, however, is to run multiple algorithms on the same network or to choose a technique that performs a local level optimization.

9.5 Local Versus Global Applications

After the presentation of the different approaches for both the creation of networks and for the most commonly used recent algorithms for community detection together with possible methodological problems, I want to tackle the problems that are introduced by the density of the underlying networks and the problems imposed on the validity of the application of these techniques in the framework of global topic detection and science mapping.

The validity of a combination of techniques does not solely depend on the accuracy of the results of the analysis. This accuracy can be measured by looking at the structure of the partition (e.g., silhouette values, [9.71]) or at consistency among different partitions (adjusted Rand-index; [9.72]). Accuracy could also be measured using additional data like lexical information or funding and grant information. But partitions (and also any other result of an analysis) can be very accurate but not relevant to the underlying question. In an earlier edition of the *Handbook of Quantitative Science and Technology Research* [9.9] *Noyons* states correctly that in science mapping exercises within a science policy context: "the actual question to be answered should be explicit". It is in the light of the actual question that the choices regarding data source, type of document link and clustering algorithm have to be made.

One issue that was only lightly touched upon in the previous sections is that of the density of the network. Density refers to the share of actual links compared to the number of possible links. With a focus on growing networks to be analyzed we are confronted with a quadratic increase of the number of possible links. Network approaches with a higher density will thus follow this quadratic growth of the number of edges when networks are extended. The analysis of such larger networks will not scale up linearly with the increase of the number of nodes but at least linearly to the number of edges.

However, none of the presented community detection algorithms discusses the scalability of the procedure with respect to the density of the network. *Waltman* and *van Eck* [9.2] for example, state that their smart local moving algorithm is capable of processing tens of millions of nodes and hundreds of millions of edges. Comparing these amounts with each other results in networks with 10 million nodes and a density of 2.0×10^{-6}. This contrasts strongly with the density of a global bibliographic coupling network. The BC-network form all citable documents in the 2013 volume of the Web of Science Core collection has a density of 0.05, which is more than 250 times higher and reaches nearly half a billion edges. Despite the enormous progress in computational capabilities and the performance of many clustering procedures, such numbers still make it very difficult to run the analysis on larger and less sparse networks.

There are different approaches to this density-imposed problem:

- Increase computational power
- Creation of large but extremely sparse networks
- Impose thresholds in weighted networks
- Improve feature extraction
- Shift the analysis to a higher level of aggregation.

One obvious solution would indeed be to *increase the computational power* one has at one's disposal. As mentioned in the introduction, services like EC2 from Amazon Web Services, Microsoft Azure and Google Cloud Platform provide very powerful data processing and analytic services at a marginal cost compared to the acquisition of similar power. But available hardware is not enough. These cloud-based services are all offering a distributed environment built on Hadoop or analytical frameworks like Spark [9.73] and not all algorithms are adapted to this Bulk Synchronous Parallel [9.74] paradigm for distributed computing. Luckily, a few volunteers have provided adapted versions of both the Louvain method (https://github.com/Sotera/spark-distributed-louvain-modularity) and the Map Equation framework (https://github.com/uwescience/GossipMap).

The next solution that has already been used in several papers [9.17, 19] is to lower the density of the network drastically by choosing the directed citation link as the connector between publications in the document network. As shown before, this will result in *extremely sparse networks*.

Waltman and *van Eck* [9.17] presented a three-level solution where the third level consists of 22 412 research areas with large differences in publications ranging from 4710 down to only 50 publications. However, this size of 50 was set in advance as a selection criterion for the inclusion of obtained clusters. The shape of the size distribution with only a few clusters with many documents and then a steep decline, followed by a long tail of small to very small clusters resembles that of cluster solutions that are very sensitive to the degeneracy problem where nearly random variations of the community structure result in similar maximization scores close to the optimal [9.70]. In contrast to the lowest level the communities from the second-level analysis (672 clusters) are plotted labeled with six major fields: mathematics and computer science; physical sciences; earth, environmental and agricultural sciences; biomedical sciences; cognitive sciences and lastly social and health sciences. This plot seems very convincing.

Klavans and *Boyack* [9.13] compared, among other classifications or taxonomies, direct citations and bibliographic coupling. Given the publication window constraints for the applicability of direct citations they used a publication window of 1996 up to 2012 while the bibliographic coupling network was restricted to 2010. Using highly cited papers as the reference for the measurement of accuracy they found that the DC approach had a slightly higher score than BC. This improved accuracy but comes with the necessary extension of the publication window.

Using a *threshold* at the level of calculated similarity or at the level of source data is also a proven strategy. As early as 1974, Small and Griffith had already applied a threshold in their cocitation-based global map of science. Their starting point was the identification of a set of highly cited papers and to calculate cocitation links among them. *Klavans* and *Boyack* [9.13] removed those references which were cited more than 100 times for the creation of their bibliographic coupling. And the Drakkar technique I developed [9.26] analyses the amount of information associated to the cited references based on the number of citations. The very large number of references with low citation rate provide a large amount of information responsible for the detection of the strong links among documents while the highly cited references are capable of connection to many papers with only low similarity.

Also for the application of lexical similarity is the application of a certain threshold on similarities a valu-

Table 9.5 Rank, label and number of journals of the top and lowest ranked clustering in a journal crosscitation network as reported by *Rosvall* and *Bergstrom* [9.63]

Rank	Label	Journals	Rank	Label	Journals
1	Molecular & cell biology	512	79	Creativity research	2
2	Medicine	765	80	Travel sociology	3
3	Physics	502	81	Medical informatics	3
4	Neuroscience	224	82	Leprosy	2
5	Ecology & evolution	349	83	Sociology (Russian)	3
6	Economics	159	84	Cryobiology	2
7	Geosciences	223	85	Death studies	2
8	Psychology	210	86	Rehabilitation counseling	2
9	Chemistry	145	87	Steel	2
10	Psychiatry	178	88	Futurist	1

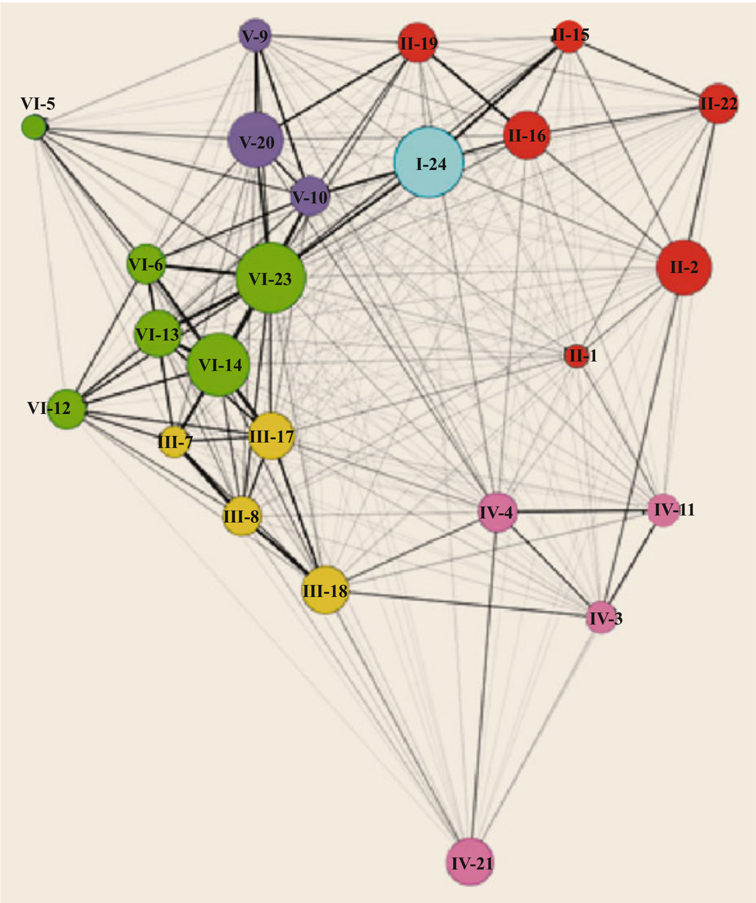


Fig. 9.4 Global map of science using 24 clusters of journals using a second-order bibliographic coupling network. Data sourced from Clarivate Analytics Web of Science Core Collection, figure reprinted from [9.75]

able approach. The threshold could be imposed in analogy to *Klavans* and *Boyack* [9.13] or *Thijs* [9.26], where the most frequently used terms or noun phrases are removed from the term list.

Using a threshold for the removal of cited references or used terms can be seen as a simple strategy for feature selection. But more advanced approaches for improved *feature extraction* or selection are available. The extraction of noun phrases [9.38] proved to be a successful strategy to reduce the density of the lexical network. But a lot of improvement can still be expected from this approach. Citation context [9.76] can be used for an adapted version of bibliographic coupling where only references cited in the same section or paragraph are considered to be linked.

Table 9.6 Codes and labels for global clustering of a journal network based on bibliographic coupling

Level I – code	Level III – code	Level III – Main subfields
I	24	Chemistry; material sciences
II	1	Statistics & probability
	2	Computer science; applied mathematics
	15	Physics; astronomy & astrophysics
	16	Engineering; classical physics
	19	Geosciences; geography
	22	Pure mathematics
III	7	Neuroscience; neurology
	8	Psychology; psychiatry
	17	Public health; nursing
	18	Social psychology; therapy; counseling
IV	3	Management; marketing; innovation
	4	Sociology; social & political sciences; law
	11	Economics; accounting
	21	Arts & humanities
V	9	Agriculture; plant science
	10	Microbiology; biotechnology; food sciences
	20	Biology
VI	5	Veterinary sciences; animal sciences
	6	Immunology; respiratory medicine
	12	Noninternal medicine
	13	Haematology; oncology; surgery; radiology
	14	Internal medicine; cardiovascular medicine
	23	Biosciences; biomedical research

Data sourced from Clarivate Analytics Web of Science Core Collection, table sourced from [9.75]

The last possible solution that is presented in this chapter to create a global map of science is to shift to a higher level of aggregation. The most natural next level of aggregation is then to cluster the scientific journals. Because of the fact that the number of nodes in the network is strongly reduced, many researchers have already produced such maps and taken different approaches. Very relevant to this chapter is the analysis by *Rosvall* and *Bergstrom* [9.7] in their paper which introduces the map equation framework. They used their flow-based approach on a journal crosscitation network with 6128 nodes and more than 6 million directed links between the journals. This resulted in 88 modules or clusters of large heterogeneity in size which they plotted in a U-shaped map with engineering and social sciences at both ends and physics, chemistry, (molecular biology) and medicine the connecting path between these ends. A short cut between both ends is provided by probability and statistics. This organization of the areas is also in line with the plot of the second-level clustering and six labels presented by *Waltman* and *van Eck* [9.17]. Table 9.5 provides the labels and the shares of the ten largest and ten smallest clusters.

While the top-ranked clusters in the left half of the table are convincing the tail with many very small clusters with only a few journals is rather peculiar. It is hard to understand that a topic like Russian sociology or Death Studies should be placed at the same level as Medicine in a global map of science. It would have been better if the final solution was limited to a set of clusters with a size above some threshold. The reported list of modules suffers from the degeneracy problem as described by *Good* et al. [9.4].

An approach that combines bibliographic coupling at the journal level with second-order similarities was taken by *Thijs* et al. [9.75]. Second-order similarities build on the concept of Friend-Of-A-Friend from social network analysis [9.77]. BC-links are calculated in a journal-by-journal matrix and, next, pairwise cosine similarities are obtained from journal vectors holding the BC-similarity to all other journals. The traditional hierarchical clustering with Ward agglomeration [9.78] was used to create clusters at three different levels. Figure 9.4 presents the structure of the network with the coloring and labeling according to the first and third level. This figure also presents a U-shaped map from engineering to social sciences with the described path between both fields and even the statistics & probability-shortcut is present.

Table 9.6 gives the codes and labels at each level. This approach resulted in 24 clusters at the third level. However, because of the very dense structure also visible in the figure (the network is almost complete) any analysis at a deeper level trying to identify smaller topics or clusters will fail due to both the resolution limit and the degeneracy problem. The high density will result in a nearly random local moving and merging from journals to clusters and will lead to arbitrary results. Also *Klavans* and *Boyack* [9.13] found similar results when comparing journal-based versus paper-based clustering at higher levels of granularity.

9.6 Conclusions

In this chapter, an overview was given of two drivers of the advancements in science mapping and topic detection, i. e., the developments in scientometrics itself and that in network science. Although the application shows an enormous constant growth in the size of the underlying networks, it becomes clear that the driver for this growth was not the advancement in scientometrics itself but rather a consequence of the opportunities offered by the increased computational resources and the advancement in scalability of community detection techniques. This chapter points the scientometrician and users of science maps to this discrepancy in growing pace of those three motors.

An important second track of research in network science relevant to topic detection is the investigation of the anomalies and artifacts created by community detection techniques based on either modularity or on the map equation (to a lesser extent). This literature clearly shows that problems like resolution limit or the degeneracy problem will always be present in any clustering approach and can place a serious burden on the validity of the obtained results. As shown in the literature, these problems start to emerge at a too fine grained level. When running the different maximization algorithms at low levels of granularity and many clusters, the formula becomes a summation of a long list of small terms. The nearly random shift or movement of nodes to another cluster results in marginal changes (increase or decrease) of these small terms. The different studies presented above all agree on a kind of U-shaped global map of science with a central role for biology and bioscience and medicine and social sciences and engineering at both ends. Any deeper clustering comes with much lower accuracy and solutions start to diverge.

This brings us to the issue of validity. As mentioned in the last section, scientometrics has always been strongly rooted in science policy and management at different levels of aggregation. In line with this applied nature of the field, the validity of the techniques suggested and used in quantitative science studies must be directly related to the research question that has to be answered. And the analysis should be performed at the most suited level or scale without a one-map-fits-all-needs approach. Science maps should not only be accurate and complete but they should in the first place be helpful to their users or readers for the cognitive structuring of their surrounding scientific reality.

References

9.1 M.E.J. Newman, M. Girvan: Finding and evaluating community structure in networks, Phys. Rev. E **69**, 026113 (2004)

9.2 L. Waltman, N.J. van Eck: A smart local moving algorithm for large-scale modularity-based community detection, Eur. Phys. J. B **86**(11), 471 (2013)

9.3 S. Fortunato, M. Barthélemy: Resolution limit in community detection, PNES **104**, 36 (2007)

9.4 B.H. Good, Y.-A. de Montojoye, A. Clauset: Performance of modularity maximization in practical contexts, Phys. Rev. E **81**, 046106 (2010)

9.5 G.E. Moore: Cramming more components onto Integrated Circuits, Electronics **38**(8), 33–35 (1965)

9.6 R. Nambiar, M. Puess: Transaction performance vs. Moore's Law: A trend analysis. In: *TPCTC 2010: Performance Evaluation, Measurement and Characterization of Complex Systems* (Springer, Berlin, Heidelberg 2011) pp. 110–120

9.7 M. Rosvall, C.T. Bergstrom: Maps of information flow reveal community structure in complex networks, PNAS **105**, 1118 (2008)

9.8 V.D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre: Fast unfolding of communities in large networks, J. Stat. Mech. Theory Exp. **2008**(10), P10008 (2008)

9.9 E.C.M. Noyons: Science maps within a science policy context. In: *Handbook of Quantitative Science and Technology Research*, ed. by H.F. Moed, W. Glänzel, U. Schmoch (Springer, Dordrecht 2004) pp. 187–213

9.10 E. Garfield: Permuterm subject index – The primordial dictionary of science, Curr. Contents **12**(22), 4 (1969)

9.11 H. Small, B.C. Griffith: The structure of scientific literatures, I: Identifying and graphing specialties, Soc. Stud. Sci. **4**, 17–40 (1974)
9.12 M.M. Kessler: Bibliographic coupling between scientific papers, Am. Doc. **14**, 10–25 (1963)
9.13 R. Klavans, K.W. Boyack: Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge?, JASIST **68**(4), 984–998 (2017)
9.14 D.M. Blei, A.Y. Ng, M.I. Jordan: Latent Dirichlet allocation, J. Mach. Learn. Res. **3**, 993–1022 (2003)
9.15 S. Wasserman, K. Faust: *Social Network Analysis: Methods and Applications* (Cambridge Univ. Press, New York 1994)
9.16 H.D. White, K.W. McCain: Visualizing a discipline: An author co-citation analysis of information science, 1972–1995, J. Am. Soc. Inf. Sci. **49**, 327–355 (1998)
9.17 L. Waltman, N.J. van Eck: A new methodology for constructing a publication-level classification system of science, JASIST **63**(12), 2378–2392 (2012)
9.18 N.J. van Eck, L. Waltman: Citation-based clustering of publications using CitNetExplorer and VOSviewer, Scientometrics **111**(2), 1053–1070 (2017)
9.19 K.W. Boyack, R. Klavans: Including non-source items in a large-scale map of science: What difference does it make?, J. Informetr. **8**, 569–580 (2014)
9.20 K.W. Boyack, R. Klavans: Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately?, JASIST **61**(12), 2389–2404 (2010)
9.21 R. Klavans, K.W. Boyack: Toward an objective, reliable and accurate method for measuring research leadership, Scientometrics **82**(3), 539–553 (2010)
9.22 E. Garfield, M.V. Malin, H. Small: Citation Data as Science Indicators. In: *Toward a Metric of Science: The Advent of Science Indicators*, ed. by Y. Elkana, J. Lederberg, R.K. Merton, A. Thackray, H. Zuckerman (John Wiley & Sons, New York 1978) pp. 179–207, reprinted in Essays of an Information Scientist, Vol. 6, p. 580, 1983
9.23 R. Klavans, K.W. Boyack: Using global mapping to create more accurate document-level maps of research fields, JASIST **62**(1), 1–18 (2011)
9.24 W. Glänzel, H.J. Czerwon: A new methodological approach to bibliographic coupling and its application to the national, regional and institutional level, Scientometrics **37**(2), 195–221 (1996)
9.25 G.M. Sheldrick: A short history of SHELX, Acta Chrystallogr. Sect. A **64**(1), 112–122 (2008)
9.26 B. Thijs: Drakkar: A graph based all-nearest neighbour search algorithm for bibliographic coupling. In: *Proc. 5th Worksh. Bibliometr.-Enhanc. Inf. Retriev. (BIR)*, Vol. 1823 (2017) pp. 101–111
9.27 M. Callon, J.P. Courtial, F. Laville: Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry, Scientometrics **22**(1), 155–205 (1991)
9.28 R.J.W. Tijssen, A.F.J. Van Raan: Mapping changes in science and technology: Bibliometric co-occurrence analysis of the R&D literature, Eval. Rev. **18**(1), 98–115 (1994)
9.29 P. Glenisson, W. Glänzel, O. Person: Combining full-text analysis and bibliometric indicators. A pilot study, Scientometrics **63**(1), 163–180 (2005)
9.30 P. Glenisson, W. Glänzel, F. Janssens, B. De Moor: Combining full text and bibliometric information in mapping scientific disciplines, Inf. Process. Manag. **41**, 1548–1572 (2005)
9.31 M. Zitt, E. Bassecoulard: Development of a method for detection and trend analysis of research fronts built by lexical or co-citation analysis, Scientometrics **30**, 333–351 (1994)
9.32 R. Todorov: Displaying content of scientific journals: A co-heading analysis, Scientometrics **23**(2), 319–334 (1992)
9.33 K.W. Boyack, D. Newman, R.J. Duhon, R. Klavans, M. Patek, J.R. Biberstine: Clustering more than two million biomedical publications: Comparing the accuracies of nine text-based similarity approaches, PLoS One **6**(3), e18029 (2011)
9.34 M.F. Porter: An algorithm for suffix stripping, Program **14**(3), 130–137 (1980)
9.35 T. Dunning: Accurate methods for the statistics of surprise and coincidence, Comput. Linguist. **19**, 61–74 (1993)
9.36 E. Leopold, M. May, G. Paaß: Data mining and text mining for S&T research. In: *Handbook of Quantitative Science and Technology Research*, ed. by H.F. Moede, W. Glänzel, U. Schmoch (Springer, Dordrecht 2004) pp. 187–213
9.37 G. Neumann, J. Piskorski: A shallow text processing core engine, Comput. Intell. **18**(3), 451–476 (2002)
9.38 B. Thijs, W. Glänzel, M. Meyer: Using noun phrases extraction for the improvement of hybrid clustering with text- and citation-based components. The example of "Information system research". In: *Proc. Worksh. Mining Sci. Papers: Comput. Linguist. Bibliometr. International Society of Scientometrics and Informetrics Conference (ISSI), Istanbul*, Vol. 1384 (2015)
9.39 G.J. Udo, R.C. Kick: The determinants of the critical success factors of information systems downsizing, Eur. J. Inf. Syst. **6**(4), 218 (1997)
9.40 W. Glänzel, B. Thijs: Using hybrid methods and 'core documents' for the representation of clusters and topics: the astronomy dataset, Scientometrics **111**(2), 1071–1087 (2017)
9.41 G. Salton, C. Buckley: Term-weighting approaches in automatic text retrieval, Inf. Process. Manag. **24**, 513–523 (1988)
9.42 B. Thijs, W. Glänzel, M. Meyer: Improved lexical similarities for hybrid clustering through the use of noun phrases extraction. In: *FEB Research Report MSI_1703* (KU Leuven – Faculty of Economics and Business, Leuven 2017)
9.43 K. Spärck Jones: A statistical interpretation of term specificity and its application in retrieval, J. Doc. **28**, 11–21 (1972)
9.44 T. Hofman: Unsupervised learning by probabilistic latent semantic analysis, Mach. Learn. **42**, 177–196 (2001)
9.45 R. Koopman, S. Wang, A. Scharnhorst: Contextualization of topics: Browsing through the universe

of bibliographic information, Scientometrics **111**(2), 1071–1087 (2017)
9.46 K. Spärck Jones, S. Walker, S.E. Robertson: A probabilistic model of information retrieval: Development and comparative experiments. Part 1, Inf. Process. Manag. **36**, 779–808 (2000)
9.47 C.D. Manning, P. Raghavan, H. Schütze: *Introduction to Information Retrieval* (Cambridge Univ. Press, Cambridge 2008)
9.48 D. Ravichandran, P.E. Pantel: Hovy: Randomized algorithms and NLP: Using locality sensitive hash function for high speed noun clustering. In: *Proc. 43rd Annu. Meet. Assoc. Comput. Linguist* (2005) pp. 622–629
9.49 J. Bichteler, E.A. Eaton: The combined use of bibliographic coupling and co-citation for document retrieval, JASIST **31**(4), 278–282 (1980)
9.50 R.R. Braam, H.F. Moed, A.F.J. van Raan: Mapping of science by combined co-citation and word analysis, part 1: Structural aspects, JASIST **42**(4), 233–251 (1991)
9.51 R.R. Braam, H.F. Moed, A.F.J. van Raan: Mapping of science by combined co-citation and word analysis part II: Dynamical aspects, JASIST **42**(4), 252–266 (1991)
9.52 F. Janssens, P. Glenisson, W. Glänzel, B. De Moor: Co-clustering approaches to integrate lexical and bibliographical information. In: *Proc. of the 10th Int. Conf. Int. Soc. Scientometr. Informetr. (ISSI)* (Karolinska Univ. Press, Stockholm 2005) pp. 284–289
9.53 R. Albert, A.-L. Barabási: Statistical mechanics of complex networks, Rev. Mod. Phys. **74**(1), 47–97 (2002)
9.54 F. Janssens, W. Glänzel, B. De Moor: A hybrid mapping of information science, Scientometr. **75**(3), 607–631 (2008)
9.55 W. Glänzel, B. Thijs: Using 'core documents' for detecting and labelling new emerging topics, Scientometrics **91**(2), 399–416 (2012)
9.56 W. Glänzel, B. Thijs: Using 'core documents' for the representation of clusters and topics, Scientometrics **88**(1), 297–309 (2011)
9.57 M.E.J. Newman: Modularity and community structure in networks, PNAS **103**(23), 8577–8582 (2006)
9.58 R.D. Bock, S.Z. Husain: An adaptation of Holzinger's B-coefficients for the analysis of sociometric data, Sociometry **13**, 146–153 (1950)
9.59 R. Rotta, A. Noack: Multilevel local search algorithms for modularity clustering, J. Exp. Algorithmics (2011), https://doi.org/10.1145/1963190.1970376
9.60 C.E. Shannon, W. Weaver: *The Mathematical Theory of Communication* (Univ. of Illinois Press, Champaign 1949)
9.61 S. Brin, L. Page: The anatomy of a large-scale hypertextual Web search engine, Comput. Netw. ISDN Syst. **30**, 107–117 (1998)
9.62 L. Bohlin, D. Edler, A. Lancichinetti, M. Rosvall: Community detection and visualization of networks with the map equation framework. In: *Measuring Scholarly Impact: Methods and Practice*, ed. by Y. Ding, R. Rousseau, D. Wolfram (Springer, Cham 2014)
9.63 M. Rosvall, C.T. Bergstrom: Mapping change in large networks, PLoS ONE **5**(1), e8694 (2010)
9.64 M.T. Schaub, R. Lambiotte, M. Barahona: Encoding dynamics for multiscale community detection: Markov time sweeping for the map equation, Phys. Rev. E **86**, 026112 (2012)
9.65 M. Kheirkhahzadeh, A. Lancichinetti, M. Rosvall: Efficient community detection of network flows for varying Markov times and bipartite networks, Phys. Rev. E **93**, 032309 (2016)
9.66 A.V. Esquivel, M. Rosvall: Compression of flow can reveal overlapping-module organization in networks, Phys. Rev. X **1**, 021025 (2011)
9.67 M. De Domenico, A. Lancichinetti, A. Arenas, M. Rosvall: Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems, Phys. Rev. X **5**, 011027 (2015)
9.68 V.A. Traag, P. Van Dooren, Y. Nesterov: Narrow scope for resolution-limit-free community detection, Phys. Rev. E **84**(1), 016114 (2011)
9.69 T. Kawamoto, M. Rosvall: Estimating the resolution limit of the map equation in community detection, Phys. Rev. E **91**, 012809 (2015)
9.70 A. Lancichinetti, S. Fortunato: Limits of modularity maximization in community detection, Phys. Rev. E **84**, 066122 (2011)
9.71 P.J. Rouseeuw: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, J. Comput. Appl. Math. **20**(1), 53–65 (1987)
9.72 N.X. Vinh, J. Epps, J. Bailey: Information theoretic measures for clustering comparison: Is a correction for chance necessary? (PDF). In: *ICML '09: Proc. 26th Annu. Int. Conf. Mach. Learn. ACM* (2009) pp. 1073–1080
9.73 M. Zaharia, M. Chowdhury, T. Das, A. Dave, J. Ma, M. McCauley, M.J. Franklin, S. Shenker, I. Stoica: Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing. In: *9th USENIX Symp. Netw. Syst. Design Implement., San Jose* (2012)
9.74 L.G. Valiant: A bridging model for parallel computing, Commun. ACM **33**(8), 103–111 (1990)
9.75 B. Thijs, L. Zhang, W. Glänzel: Bibliographic coupling and hierarchical clustering for the validation and improvement of subject-classification schemes, Scientometrics **105**(3), 1453–1467 (2015)
9.76 K.W. McCain, K. Turner: Citation context analysis and aging patterns of journal articles in molecular genetics, Scientometrics **17**(1), 127–163 (1989)
9.77 B. Thijs, E. Schiebel, W. Glänzel: Do second-order similarities provide added-value in a hybrid approach?, Scientometrics **96**(3), 667–677 (2013)
9.78 J.H. Ward Jr.: Hierarchical grouping to optimize an objective function, J. Am. Stat. Assoc. **58**, 236–244 (1963)

Bart Thijs

ECOOM
KU Leuven
Leuven, Belgium
bart.thijs@kuleuven.be

Bart Thijs is a Senior Researcher in Bibliometrics at KU Leuven – ECOOM. His research focusses on the identification of global and local network structures within scientific communication and on the development and application of citation-based indicators. He combines a background from Mathematical Psychology (Master's degree from KU Leuven) and Social Sciences (PhD from Leiden University) with professional experience as a statistical consultant in his work.

Part B Advancement

Part B Advancement of Methodology for Research Assessment

10. Measuring Science: Basic Principles and Application of Advanced Bibliometrics

Anthony van Raan

We begin with a short history of measuring science and discuss how the Science Citation Index has revolutionized the quantitative study of science and created a strong application potential. After reviewing the rationale of bibliometric analysis, we present the basic principle of the bibliometric methodology, with complex citation networks as a starting point. We show that the two main pillars of advanced bibliometric methods, citation-based analysis and science mapping, are both reducible to one and the same principle. From this basic principle we deduce a set of main indicators, particularly for the assessment of research output and international impact. Important elements include new approaches for identifying fields and research themes on the basis of a publication-level rather than a journal-level network; publication and citation counting; normalization of citation measures; the use of indicators based on averages versus those based on citation distributions; and weighting procedures and statistical reliability. In this account of the state of the art of advanced bibliometrics, we highlight in particular the developments in our Leiden institute, given its long-standing, extensive, and broad experience.

The next part of this chapter deals with practical applications of indicators, particularly real-life examples of evaluation studies. We further discuss several crucial issues such as the use of journal impact factors and h-index; the relation between peer review judgment and bibliometric findings; definition and delimitation of fields; assignment of publications; the influence of open access; webometrics and altmetrics; ranking of universities; and general objections to bibliometric analysis.

The second main pillar of the advanced bibliometric methodology is the development of science maps. We discuss the basic elements and the construction of both citation-relation and word-relation science maps. Further, we present a method to combine the two main pillars: the integration of citation analysis in science maps. This combined citation analysis and science mapping can be used to explore research related to socioeconomic problems. Recently developed bibliometric instruments enable tunable mapping, which opens up new analytical opportunities in monitoring scientific research. Finally, we contend that bibliometric indicators and maps are not just evaluation tools for science policymakers, research managers, and individual researchers, but also powerful instruments in the study of science.

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_10

10.1 A Short History of Scientometrics

The quantitative study of science, generally referred to as *scientometrics*, is aimed at advancing our knowledge on the development of science and its communication structure, as well as in relation to social, technological, and socioeconomic aspects. This field is both problem-oriented and basic in nature. There are important interdisciplinary links with the philosophy, history and sociology of science, with policy and management studies, mathematics and physics, with innovation studies, and particularly with information science. Within scientometrics, the research on scientific communication, in particular with data from publications, citations, and journals, is called *bibliometrics*. For more compact reviews we refer to earlier papers of the author [10.1–3] on which this handbook chapter is partly based and from which pieces of text that can be regarded as basic descriptions are taken. An impressively comprehensive overview of the early developments in the quantitative study of science was given in 1976 by *Francis Narin* in his seminal report, *Evaluative Bibliometrics* [10.4].

10.1.1 The Quantitative Study of Science Before the Science Citation Index

Perhaps the first scientometric study was the work in 1873 by *Alphonse de Candolle* [10.5], in which he described the changes in the scientific strength of nations by membership of scientific societies. He tried to find *environmental factors* of all kinds, but particularly religion (Protestant versus Roman Catholic, including the role of the celibate), for the scientific success of a nation. In 1926, *Alfred Lotka*—famous for his Lotka–Volterra equations in population dynamics—published a remarkable study on the productivity of chemistry researchers [10.6]. It is the first scientometric finding that can be expressed in a simple mathematical way: the productivity of scientists follows an inverse square law. In other words, the number of scientists producing N papers is proportional to N^{-2}. This *law of scientific productivity* holds that within a given period of time, the majority of researchers produce only one or two papers, and just a few authors produce ten or more papers. It is the first evidence, later followed by much more, that science is characterized by skewed distributions. Another early (1940s) discovery of skewed distribution in science was that the literature on a given subject was heavily concentrated in just a small core set of journals [10.7]. This phenomenon, known as *Bradford's law* provided a convenient way to estimate the number of journals that must be checked to obtain a specific degree of coverage of the literature on the subject.

Between 1940 and 1950, we see increasing interest on the part of librarians in determining how papers in specific journals are cited by papers in other journals. The idea emerged that analyzing these journal-to-journal citations could be used to assess the usefulness of journals for university libraries. This works as follows [10.4]: Given a set of journals already present in a library, one may analyze which journals outside the set are frequently cited by the journals within the set. These journals can then be considered important enough to add to the library collection. The beginning of the 1950s witnessed the appearance of the first *science maps* based on journal-to-journal citations. *Robert Daniel* and *Chauncey Louttit* constructed such a map for the field of psychology in order to show its development based on structures hidden in the literature [10.8]. By composing a journal-to-journal-citation matrix, they calculated similarity measures and applied cluster analysis. In this way they created a science map (given in [10.4]) and discovered a general and an applied core of the psychology literature. All this had to be done by hand, with a huge amount of printed journal volumes, a hell of a job without computers and automated databases …

Initially, citation analysis was restricted to journals as a whole. Citation analysis of individual publications, grouped into fields, countries, universities, departments, research groups, or individual scientists, was not possible due to the lack of relevant data sources. All such studies had to wait for the creation of the *Science Citation Index* in 1955.

10.1.2 The Science Citation Index Revolutionized the Study of Science

The creation of the *Science Citation Index* (SCI) by *Eugene Garfield* [10.9] was undeniably a major breakthrough [10.10], which completely revolutionized the field. The systematic collection and registration of all references in all publications in a large number of scientific journals opened the way to statistical analyses of the scientific literature on a very large scale. It marks the rise of bibliometrics as an application-driven empirical field within the studies of science. Garfield founded a company, the *Institute for Scientific Information* (ISI), to develop the SCI and related databases. The SCI is the predecessor of the *Web of Science* (WoS). As most publications are connected to other publications by citations, the science communication system represented by publications is in fact a gigantic network to which currently around two million publications per year are added. It is therefore no surprise that in bibliometric research today, advanced network analysis methods from mathematics and physics are used.

Eminent US scientists such as *Derek de Solla Price* [10.11, 12] and *Robert Merton* [10.13] immediately recognized the enormous potential of Garfield's creation: Price, from the perspective of contemporaneous history of science, and Merton from the perspective of normative sociology. Price is the pioneer of a *physical approach* to science, wherein he tried to find laws to predict further developments, inspired by the ideas of Newtonian and statistical mechanics. In this approach, quantitative measures of science, *indicators*, are guides to finding and, as a crucial next step, understanding basic features of the scientific communication system. He made a number of important bibliometric discoveries based on SCI data: *Price's law* on the distribution of publications over authors, stating that 25% of all authors are responsible for 75% of all published papers (this is in fact a more accurate revival of the Lotka study); the exponential growth of scientific literature that started at the beginning of the 18th century; the exponential decay of citations received by publications as a function of time, which defines a *half-life* of publications; and the *power law* distribution of citations over publications. This last finding is the most important example of skewed distributions in science: the vast majority of publications receive only a few citations (or none), and only a few publications receive many citations [10.11, 12, 14]. *Price* also proposed the idea of *cumulative advantage* [10.15], which means that publications which are already frequently cited have a higher probability of receiving even more citations. In complex network theory, this concept is now better known as *preferential attachment*.

The creation of the SCI opened the way for the analysis of citations at the level of individual publications. In the early 1960s, *Michael Kessler* at MIT developed the method of *bibliographic coupling* (BC): two publications are bibliographically coupled if they have references in common; the more common references they have, the stronger their relation [10.16]. Ten years later, Henry Small at ISI developed the mirror of bibliographic coupling, which he named *co-citation analysis* (CC). With this method, publications are defined as related if they are cited together by other papers: the more papers cite a specific pair of papers (references), the stronger the co-citation strength of these cited papers. The strength of the relations between publications provides similarity measures, and hence the possibility to cluster in such a way that both BC and CC can be used for mapping [10.17]. In Sect. 10.2 we will discuss these bibliometric methods in detail.

The SCI also made it possible to use citation analysis to assess the impact of publications: the more they are cited, the higher the impact. From the beginning it was recognized that the process of citation is complex, and certainly does not provide an ideal monitor of scientific performance. This is particularly the case at a statistically low aggregation level, such as an individual publication. But the application of citation analysis to the work of a group of researchers as a whole over a longer period of time does yield (particularly in the natural sciences and the medical fields) a reliable indicator of scientific performance, as will be discussed in this chapter. The sociologists Jonathan Cole and Stephen Cole were the first to use citation analysis to determine scientific impact. They found high positive correlations between receiving citations, winning awards, membership in scientific academies, being widely known among colleagues, and working in prestigious institutes. They also concluded that it seemed as though only a small number of scientists contribute to scientific progress [10.18, 19].

The 1970s saw a rapid increase in research on quantitative aspects of science. Due to the increasing availability of data, the quantitative appraisal of science gained influence in national and international organizations as well. It was the beginning of the development of *science indicators* based on publication and citation data. Stimulated by the success achieved by economists in developing quantitative measures of political significance (e. g., unemployment, gross national product), the National Science Foundation, UNESCO, the OECD, and the European Commission (EC) began systematically collecting data to measure and analyze the development of science and technology. A landmark publication was the first report of a biennial series of *Science Indicators* published by the US National Science Board in 1973 [10.20]. We note that the first OECD report on *The Measurement of Scientific and Technological Activities*, the famous Frascati manual, had appeared already 10 years earlier, in 1963 [10.21].

Meanwhile, the creator of the SCI, *Garfield*, introduced the concept of a *journal impact factor*, in which he used citation analysis as a tool for journal evaluation [10.22, 23]. Here, Garfield recognized another potential application of citation analysis: journals can be ranked by citation frequency for science policy purposes. The policymaking implications of bibliometric analyses were discussed even more explicitly in the pioneering work by *Narin* on evaluative bibliometrics mentioned earlier [10.4]. Narin was the first scientist-entrepreneur who used SCI data for a commercial enterprise, in a company called Computer Horizons, Inc. With a research contract from the National Science Foundation, Narin extensively described the use of publication and citation analysis in the evaluation of scientific activities. The focus was on creating performance indicators based on the data of thousands of journals. He also developed a new journal impact indicator utilizing a recursive method, the *journal influence weight* [10.24]. In recursive indicators, citations receive a higher weight the more their citing sources themselves are cited. *Narin* also wisely realized that patent analysis was necessary to explore the links between science and technology, and that a combination of citation and patent databases would be crucial [10.25]. An analysis on the basis of millions of individual publications was at that time still a bridge too far. However, in addition to the availability of more and more data through the SCI, computer power and memory capacity were rapidly increasing (by the end of the 1990s, computing power and memory storage capacity was about 10 000 times what it was in the early 1980s; nowadays it is about 10 000 000 times). It took only a few more years to bring about large-scale citation analysis of (sets of) individual publications.

10.1.3 Europeanization and Further Development of Bibliometrics

Until that time, there was not been much activity in bibliometrics outside the United States, but this changed around 1975. During a 4-month working visit at the National Science Foundation, Cees le Pair, the then director of research at FOM (*Fundamenteel Onderzoek der Materie*, the physics research council, part of the Netherlands Organization for Scientific Research), became interested in *measuring science* and got to know Derek de Solla Price and Eugene Garfield. He conducted the first citation analysis studies of two fields of physics: magnetic resonance and electron microscopy [10.26, 27]. These studies showed that citation analysis made it possible to identify the most important basic research contributions to these fields. In the field of magnetic resonance, high correlations were found between the results of citation analysis and peer assessments. However, in the case of applied and technological research—particularly in the field of electron microscopy—citation analysis did not work that well. Patent analysis appeared to be necessary to adequately describe the developments. Clearly, the limits of the bibliometric methodology were already understood in these early studies.

Around the same time, the Hungarian chemist Tibor Braun established a flourishing bibliometric research unit at the Hungarian Academy of Sciences in Budapest. At first, the results of bibliometric research were published mainly in sociology or general journals, even top journals such as *Science*. An important sign of the emancipation of the field of bibliometrics was the creation of its first journal, *Scientometrics*, by Braun in 1978. *Scientometrics* is still one of the core journals of the field. In the same year, 1978, another milestone was reached: the publication of *Toward a Metric of Science: The Advent of Science Indicators* [10.28]. Many issues discussed in this remarkable book are still relevant today. In the 1970s, important issues in bibliometric research included the use of co-citation analysis to structure fields of science, automatic indexing of search terms in databases, journal interrelationships, scientific collaboration, the role of science in innovation, the growth of science, mobility and scientific careers [10.29–32].

In 1979 the Leiden University executive board introduced a new institutional policy: the funds allocation model used to divide the block financing from the Ministry among faculties should no longer be primarily dependent on student numbers, but should also have a research quality-dependent factor. The crux was to find a reliable method for assessing research quality. In 1980, the first experiments were performed: a citation analysis of all chemistry departments, followed by a second citation analysis study of the entire Faculty of Sciences and the Faculty of Medicine, combined with expert interviews [10.33]. Never before had a bibliometric study been performed on so many (about 140) departments within a research-intensive university. But certainly, shortly before and around the same time, there were other important bibliometric studies, of which the work on radio astronomy by *Martin* and *Irvine* in 1983 [10.34] can be considered groundbreaking.

The publishing company Elsevier became particularly interested in creating new performance indicators for and mapping of scientific journals. In the early 1980s, there was a rapid rise in the number of co-citation analysis studies including author co-citation [10.35–37]; an increasing emphasis on important themes such as advanced statistical analyses of bibliometric parameters [10.38, 39]; citation analysis in evaluation studies [10.33, 34]; application of bibliometric methods in the social sciences [10.40], comparison of peer opinions and bibliometric indicators [10.41], the development of indicators of interdisciplinary research [10.42], scientific collaboration [10.43], the normalization of citation impact [10.33], online searching [10.44], and patent indicators [10.25, 45, 46]. A new development was *co-word analysis* [10.47, 48]. Mathematically, this is similar to co-citation analysis, but instead of using the references mentioned in publications, it is based on the use of the concepts mentioned in publications. To put it simply, next to the list of references in publications, we can also characterize publications by a list of concepts used in them. After a slow start, co-word analysis skyrocketed, becoming one of the primary methods for creating science maps. But it took a long time, almost two decades, before sufficient computer power was available to create maps in a relatively short length of time so that it became possible to analyze the effects of changing the map parameters such as a threshold for the similarity strength.

The first international conference on bibliometrics and the theoretical aspects of information retrieval (the predecessor of the International Society for Scientometrics and Informetrics (ISSI) conferences) was organized in 1987 by Leo Egghe in Hasselt (Belgium), again an important step in the emancipation of the field. In 1988, the first *Handbook of Quantitative Studies of Science and Technology* was published [10.49], and the Centre for Science and Technology Studies (CWTS) organized the first International Conference on Science and Technology Indicators, in Leiden (the Netherlands). The field of bibliometric research became increasingly dynamic—and even heated. As could be expected, opponents of the use of citation analysis started to agitate the field, for instance by pointing to the many different motives authors may have to cite or not to cite, thus questioning the validity of using citations as a performance measure. Although such debates can be quite irritating, they are part of a healthy development of the field. At the same time, more and more large-scale empirical work was done. For instance, the Budapest researchers *Braun* et al. published an extensive study on the normalization of citation impact and the publication output and field-specific citation impact of scientific work in about 100 countries [10.50, 51]. Important work in the second half of the 1980s focused on areas such as research performance in the humanities and social sciences, further development of multidimensional scaling analysis for science mapping based on co-citation and co-word analysis, improvement of journal-based field classification, citer motivations and citation behavior, and interdisciplinary research.

The 1990s witnessed an increase in contract work on the use of bibliometric indicators in research evaluation, commissioned by organizations and institutions worldwide, but especially in the European Union (ministries, national research councils, charities, universities). These practical applications of bibliometric methods were extremely important: they led directly to the improvement of bibliometric indicators and maps, but also to a strong intensification of basic research. We mention work on the combination of co-citation and word analysis, the skewness of science, journal impact measures and their appropriateness in evaluation studies, the correlation between peer judgments and outcomes of citation analysis, international collaboration of the upcoming Asian powers, technology mapping based on patent analysis, and the interface of science and technology. Also, quite exotic issues such as the fractal structure of science became the subject of investigation. Furthermore, the first webometrics studies appeared [10.52–59].

In 1992, Garfield decided to sell his Institute for Scientific Information. After a few years of different owners, ISI became part of the information giant *Thomson Reuters*. Things began to change gradually, with research becoming increasingly determined by business interests. As Tibor Braun once noted, this was the end of the romantic period for the bibliometric world. In the second half of the 1990s in particular, bibliometric methods for assessing and monitoring research performance were applied on a large scale in the Netherlands. Two major programs stand out: first, the *Netherlands Observatory of Science and Technology* (NOWT), with the aim of compiling the biannual Science and Technology Indicators Report for the Ministry of Education, Culture and Sciences; and second, the VSNU (*Vereniging van Samenwerkende Nederlandse Universiteiten*, Association of Universities in the Netherlands), with detailed national research assessment procedures for disciplines such as biology, chemistry, physics, and psychology, in which extensive bibliometric analyses were performed for all research groups of these disciplines in the Netherlands. Parallel to these applications we see that bibliometric researchers strongly warn about the danger of *easy bibliometrics*, particularly the inappropriateness of journal impact factors for research evaluation [10.60–62].

10.1.4 Bibliometrics in the Internet Age

At the beginning of the new century, several significant events took place. First, not surprisingly, the Internet had changed scientific communication. In addition to the publication and citation data provided by the WoS database, a vast amount of other publication data included in institutional and personal websites became available. Thus, next to citation analysis, the use of data provided via the Internet, *webometrics*, was (and still is) considered to offer interesting opportunities to complement citation-based analysis in evaluation and mapping approaches. We discuss these developments in Sect. 10.3. An important second event was the emergence of university rankings. Probably the first ranking of European universities was the EC-commissioned study on scientific excellence within the framework of the European Research Area [10.63]. In a global context, the Academic Ranking of World Universities (ARWU, ShanghaiRanking) [10.64] was the first in 2003. The Times Higher Education (THE) launched its ranking in 2004 [10.65], and in 2007 the first Leiden Ranking [10.66, 67] was published, followed by the QS ranking (2009) [10.68], the Scimago ranking (2009) [10.69], and U-Multirank (2011) [10.70]. Despite the many problems inherent in the ranking of universities [10.71], this new phenomenon evoked a rapidly increasing public and political interest in the performance of universities. In fact, it drove the application of research assessment, particularly with bibliometric methods.

The third important event was Elsevier's development and 2004 launch of *Scopus*, the first competitor of Thomson Reuters' WoS. After a run-up of a few years, the commercial launch of Scopus as the new citation index marked the end of a very long period of strict monopoly of the WoS. A fourth remarkable event was the introduction in 2005 by *Jorge Hirsch* of the h-index [10.72]. This new indicator attracted enormous attention. A simple method for individual scientists to find their h-index is to rank their publications, for instance in the WoS, according to the number of times the publications are cited (starting with the highest-cited). Somewhere in this ranking there will be a publication with a number of citations that is the same as its rank number; that number is the h-index. A real torrent of publications describing numerous variants of the h-index followed, as we shall encounter in Sect. 10.3.

The Internet became more and more important, also for citation counting, particularly by the launch of Google Scholar in 2004 with a citation index including publications from sources in addition to journals, such as books and conferences [10.73, 74]. In the beginning of the first decade of the new century, important research issues included the triple helix model of government–industry–academy interaction [10.75]; h-index and variants; the influence of language on the citation impact of particularly German and French research groups; webometrics; university rankings, their influence and limitations; patent citation analysis and the identification of industrially relevant science; the rapidly advancing scientific performance of China; improvements in text mining; the identification of scientific excellence; and sleeping beauties (a sleeping beauty in science is a publication that goes unnoticed (i. e., is not cited, *sleeps*) for a long time and then, almost overnight, attracts a great deal of attention (*is awakened by a prince*), as measured by citations.

The years 2006–2010 can be characterized as a period of strong growth in business-based bibliometrics: commercial database producers began introducing their own bibliometric products: Elsevier's Scopus with SciVal and Thomson Reuters with InCites. In research we see a strong focus on webometrics, a further discussion of the properties of the h-index, new methods for identifying emerging topics, improvements in the visualization of science maps such as the VOSviewer, applications of Google's PageRank algorithm, source-normalized citation impact, and measures of journal interdisciplinarity. In the current decade, the number of publications on webometrics and altmetrics issues increases substantially, and the bibliometric community enters into fierce debates on new *crown indicators*, university rankings, fractional or full counting, and publication-level versus journal-level field classifications. A growing number of advanced network analytical methods are applied to bibliometric data and used to improve science maps. In Sect. 10.3 we will discuss these developments in greater detail, along with reference to relevant literature.

And so we arrive at the present-day developments in bibliometrics, and we conclude our historical overview. In the following survey of the state of the art of advanced bibliometrics, we highlight in particular the developments in our Leiden institute, given its long-standing, extensive, and broad experience. The remainder of the chapter is structured as follows. In Sect. 10.2 we discuss the rationale of and the practical needs for advanced bibliometrics. The majority of Sect. 10.2 is devoted to a discussion of the basic elements of advanced bibliometrics. Section 10.3 addresses major issues in the practical application of bibliometric performance indicators and provides a real-life example of a bibliometric evaluation study. In Sect. 10.4 we focus on the basic elements, construction, and application of science maps. We conclude the chapter in Sect. 10.5 with a few thoughts about the measurability of science.

10.2 Bibliometric Analysis: Rationale, Practical Needs, Basics

10.2.1 Why Bibliometric Analysis?

Science is an abstract concept. How can we approximate its characteristics quantitatively? In other words, how can we measure science; what is the metric? Measuring basically means counting of standard units. Measurement in physics is a well-understood process in which observed quantities are related to well-defined standard units. But in the non-physics world there are no perfect standard units, such as one hydrogen atom, one kilogram, or one meter. There are only units with a reasonable similarity, e. g., inhabitants of a city, students at a university, and publications in a field of science.

The daily practice of scientific research shows that inspired scientists, particularly in the natural sciences and medical research fields, go for publication in international journals. Certainly, journal articles are not in all fields the main carrier of scientific knowledge. And they differ widely in importance. But work of at least some importance provokes reactions from colleagues. They are the *invisible college* by which research results are discussed, and they play their role as members of this invisible college by referring in their own work to earlier work of other scientists. Scientific performance relates to the quality of the contribution in terms of *increasing our knowledge* (scientific progress) as perceived by other knowledgeable researchers (peer review), quantified and archived by citations. This process of citation is undoubtedly a complex one, and it does not provide an ideal monitor of scientific performance. This is particularly the case at a statistically low aggregation level, such as the individual researcher. But the application of citation analysis to the work of a group of researchers as a whole over a longer period of time does yield, in many situations, a strong indicator of scientific performance [10.1]. In this context, citations are seen as providing a measure of international impact, which in turn is considered a good proxy for scientific quality, also in terms of relevance and visibility. For many years, the Science Citation Index, now the Web of Science (WoS, produced by Thomson Reuters, now owned by Clarivate Analytics) was the only large multidisciplinary citation data source worldwide. In the last 10 years or so, Scopus, produced by Elsevier, has provided a second comprehensive citation database. We note that, since the early 1990s, the SLAC National Accelerator Laboratory (formerly Stanford Linear Accelerator Center) has maintained a website with a freely accessible high-energy physics database, SPIRES, including citations.

The motives for giving (or not giving) a reference to a particular article may vary considerably [10.76–82]. There is, however, no empirical evidence that these motives are so disparate or so randomly given that the phenomenon of citation would lose its role as a reliable measure of impact [10.81]. Nevertheless, contentious discussion around how authors choose their references regularly flares up, often in relation to processes of cumulative advantage such as the Matthew effect [10.13], for instance, in which authors tend to cite papers that are already highly cited [10.82]. In general, this Matthew effect is taken for granted. However, studies on PhD graduations with honors question the predominant role of the Matthew effect in acquiring citations [10.83, 84].

Why bibliometric analysis of research performance? Peer review is and must remain the principal procedure of quality judgment. But peer review is not sacrosanct. It may have serious shortcomings and disadvantages; it might even suppress innovation [10.85]. Subjectivity, i. e., dependence of the outcomes on the choice of individual committee members, is one of the major problems. This dependence may result in conflicts of interest, unawareness of quality, or negative bias against younger people or newcomers (nepotism) to the field and women (sexism) [10.86–88]. To make peer review more objective and transparent, it should be supported by advanced bibliometric methods. Of course, this does not provide an ideal instrument, working perfectly in all fields under all circumstances. But it works very well in the large majority of the natural, medical, applied, and behavioral sciences. These fields of science are the most cost-intensive and the ones with the strongest socioeconomic impact. For a comprehensive discussion on the validity and reliability of evaluation of scholarly performance, we refer to [10.89].

A first and good indication of whether bibliometric analysis is applicable to a specific field is given by the publication characteristics of the field, in particular the role of international refereed journals. If international journals are a major means of communication in a field, then in most cases bibliometric analysis is applicable. This is generally the case for the natural and medical sciences but less so for engineering, social sciences, and particularly the humanities [10.40, 90–96]. Therefore it is important to study the publication practices of a research group, department, or institute, in order to determine whether bibliometric analysis can be applied reliably. A practical measure to this end is the share of publications covered by WoS or by Scopus in the total research output. For publications not covered by the WoS or Scopus, a restricted type of analysis is possible, insofar as these publications are cited by articles in journals covered by the WoS or Scopus. This *non-source* approach is particularly important for bibliometric analysis in the social sciences and humanities [10.93, 95].

A frequently posed question concerns the delay problem: does citation analysis suffer from a substantial, *inherent delay* in the measurement of research performance [10.97]? This supposed inherent delay implies that colleague-scientists (peers) in some field of science are already familiar with a specific piece of research work, but it takes some time before they begin citing the work. In other words, a time delay between *peer awareness* and *bibliometric notification* (i. e., by citing the work). An answer to this question needs further refinement: delay compared to what? To the average processing time of a publication? To the average running time of a project? Or to peer review time cycles? The average duration of a major research project is about 4 years, and the same is true for most peer review time cycles. Also, during the publication process, the awareness of scientific community evolves (e. g., average time between field-specific conferences). Nevertheless, it is not unusual for important papers to be cited already in the year of publication. Apparently bibliometric notification does not necessarily take more time than peer awareness. But it is not that easy; circular reasoning lies in wait. How do *you* know that researchers in some field of science already know specific work? Often bibliometric notification is taken as a sign of this peer awareness—for instance, in the study of *delayed recognition*: in relatively rare cases, it takes a long time for peer awareness, reflected by the absence of citations over a longer period of years [10.98, 99]. The publications suffering from delayed recognition are called *sleeping beauties*; we come back to this phenomenon in Sect. 10.3. Perhaps altmetric methods will enable us to ascertain whether there is a time lapse between peer awareness and bibliometric notification.

As we discussed in our historical overview, the Internet has changed scientific communication. Researchers use the web for both information-seeking and presenting. In addition to the publications not covered by the WoS or by Scopus, there is a large number of other publications and data included in institutional and personal websites. As we will discuss in Sect. 10.2, data provided via the Internet, *webometrics*, offer interesting additional opportunities to aid citation-based analysis in evaluation and mapping studies.

10.2.2 Advanced Bibliometrics and Practical Needs

We define *advanced bibliometrics* as the state-of-the-art design and application of mathematically sound quantitative methods based on publication, citation, and textual data for research impact assessment and for mapping of scientific fields. Also, these methods can be applied more generally in the study of a wide range of properties of science and of its communication system, such as growth and differentiation of scientific research, diffusion of knowledge, mobility of researchers, scientific collaboration, and the identification of breakthroughs. In this chapter we focus on research impact assessment and mapping. Advanced bibliometrics is built on three cornerstones:

1. *Reliable statistics*, e. g., corrections for differences in publication and citation practices between scientific disciplines; robustness with respect to outliers
2. *Highest possible accuracy of data*, e. g., carefully cleaned institutional addresses and author names
3. *Sets of consistent*, sufficiently divers, and transparently designed and calculated quantitative measures, called *indicators*.

Above all, the most fundamental bibliometric operation, the identification and storage of *citing-cited pairs* (links between a citing and a cited paper), must be completely transparent. There can be no advanced bibliometrics if somehow and somewhere in whatever data manipulation black boxes are present. Advanced bibliometrics enables universities and research councils to support their decisions on the key questions they are facing. These are summarized in Scheme 10.1.

Scheme 10.1 Key questions of advanced bibliometrics

- What is our output and international impact, and what are the trends?
- Do we see notable changes?
- Can we identify breakthrough work?
- Where is our research located on the map of science?
- Where does our impact come from (specific institutions, fields)?
- With which universities and research institutes do our scientists cooperate, and how intensively?
- How do we distinguish ourselves to attract excellent staff and students?
- How should we organize and monitor our research given new and often interdisciplinary developments?
- How should we divide funds?
- What is the societal impact of our research?
- How do we rate output and quality compared to input and composition of a research unit?
- Can we apply bibliometric indicators in engineering, social sciences, and humanities?
- To what extent is our research related to new, emerging research themes?
- How can we identify experts and panel members for peer review and advice?

In order to achieve the best possible level of utilization, advanced bibliometrics must enable users to perform analyses at different aggregation levels (university, institutes, departments, groups, programs) within time periods as desired, and to create their own reports with visualizations (tables, figures, maps) of the outcomes of the analyses. In this way the users are provided with a *bibliometric monitor*. Such a monitor enables comparisons between departments and research groups within an institute, but also inter-institutional comparisons in order to identify strengths and collaboration opportunities. Of course, nothing in nature is ideal, and advanced bibliometrics also has its limitations. In the following sections we will discuss the basic elements and applications of advanced bibliometrics. We will see that there are many challenges in the application of bibliometrics in evaluation practice, that there is still a need for further improvement, and that there are things we have to live with. Making quantitative measures of anything thinkable fascinates many of people, but it horrifies others as being nonsense and taking us back to a magic number world. Because scientific quality is so precious and often regarded as immeasurable in all its aspects, it is no wonder that the application of advanced bibliometrics may evoke strong emotional reactions, as we will see in this chapter.

10.2.3 The Fundament of Bibliometric Methods: The Publication-Attribute Network

The fundament of bibliometric analysis is the publication-attribute network (PAN). The two main methods—citation analysis and science mapping—can both be derived from the same network principle. In this approach, publications have specific *attributes*, and these relations can be represented by unidirectional linkages in the PAN, which can be regarded as the primary network. These attributes can be *cited papers*, i. e., the references of a paper, or *concepts* (keywords), as well as authors and institutions. In Fig. 10.1 we give a simple example of a PAN. We have four publications $p_1, \ldots, p_4$ and five attributes $a_1, \ldots, a_5$. Publication p_1 has all five attributes; p_4 has none. From the PAN, two secondary networks can be derived: the attribute network (AN) and the publication network (PN). The connecting lines indicate the strength of the relation between attributes or between publications. For instance, in the AN, a_1 and a_4 are connected with linkage strength 3, because these attributes have three publications in common, namely p_1, p_2, and p_3. In the PN, p_1 and p_3 are connected with linkage strength 2, because these publications have two attributes in common, namely a_1 and a_4. In Scheme 10.2, we give a simple mathematical explanation of the relation between the primary and the two secondary networks.

If the attributes are *cited papers* (references), the PAN represents *direct citation* (DC) of the cited papers (as attributes) by the citing papers. The AN represents the co-citation (CC) network and the PN represents the bibliographic coupling (BC) network. In Scheme 10.2 we explain how attributes (for instance, cited papers) and publications are linked together with a strength that can be calculated.

These strengths provide similarity measures (the stronger the linkage between attributes or publications, the more similar these attributes or publications are). This enables us to make maps in which distances between attributes or publications are inversely proportional to their linkage strengths. Linkage strengths also enable us to analyze clustering of groups of attributes or publications. With the BC network, we can create maps on the basis of publications in their *citing* modality; in the CC method, the maps are constructed on the basis of their *cited* modality. As the citing modality can no longer be changed (the references in publications are fixed and thus remain forever the same), the BC maps are static, whereas the CC maps are dynamic (publications can be cited later on, again and again). For pioneering research on co-citation mapping, we refer to the work of *Small* [10.17, 100–103], and on the thematic identification of co-citation clusters to [10.52, 53].

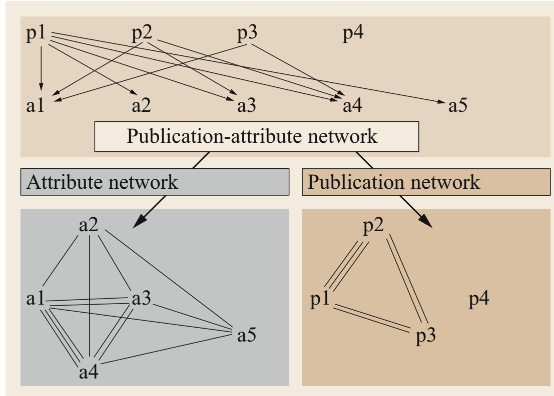


Fig. 10.1 The primary (PAN) and secondary networks (AN and PN)

Scheme 10.2 Simple mathematical explanation of the relation between the primary and the two secondary networks

In Fig. 10.1 we showed a simple PAN in which we have four publications p_1, p_2, p_3, and p_4, and five attributes a_1, a_2, a_3, a_4, and a_5. We see that p_1 contains all attributes a_1 through a_5; p_2 contains a_1, a_3, and a_4; p_3 has only a_1 and a_4; and p_4 has none of the five attributes. Mathematically, the publications can be written as vectors in the attributes space, e. g., $p_1 = (1, 1, 1, 1, 1)$. The PAN can thus be written as matrix $\mathbf{N}$ in which the rows represent the publications and the columns the attributes,

$$\begin{array}{cccccc} & a_1 & a_2 & a_3 & a_4 & a_5 \\ p_1 & 1 & 1 & 1 & 1 & 1 \\ p_2 & 1 & 0 & 1 & 1 & 0 \\ p_3 & 1 & 0 & 0 & 1 & 0 \\ p_4 & 0 & 0 & 0 & 0 & 0 \end{array} .$$

Formally written as

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} = \mathbf{N} \tag{10.1}$$

Pre-multiplication of matrix $\mathbf{N}$ with its transpose $\mathbf{N}^{\mathrm{T}}$ yields the (symmetrical) attribute-correlation matrix

$$\begin{aligned} \mathbf{N}^{\mathrm{T}} \cdot \mathbf{N} &= \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{3} & 1 & 2 & 3 & 1 \\ 1 & \mathbf{1} & 1 & 1 & 1 \\ 2 & 1 & \mathbf{2} & 2 & 1 \\ \mathbf{3} & 1 & 2 & \mathbf{3} & 1 \\ 1 & 1 & 1 & 1 & \mathbf{1} \end{pmatrix} = \mathbf{A} \end{aligned} \tag{10.2}$$

In this way, we have transferred the PAN (represented by matrix $\mathbf{N}$) into the AN (represented by matrix $\mathbf{A}$). The *diagonal* values (in boldface) of this matrix $\mathbf{A}$, $a(i,i)$, indicate the number of publications with attribute a_i. For a_1 this is $a(1,1) = 3$, for a_2 we find $a(2,2) = 1$, and so on. Thus, the matrix diagonal represents the *occurrence* of each attribute. The *off-diagonal* values, $a(i,j)$, give the *co-occurrences*; for instance, a_1 and a_4 (value in red) are both present in $a(1,4) = 3$ publications (namely in p_1, p_2, and p_3, as can be seen in Fig. 10.1). Thus, $\mathbf{A}$ provides the *strengths* of the links between each possible attribute pair based on the number of publications in which an attributes pair occurs. In the case the attributes are cited papers, the diagonal values are the number of times a specific paper is cited within the given set of publications; the off-diagonal values give the (not-normalized) co-citation strengths.

If we now take the mirrored matrix multiplication of (10.1), i. e., post-multiplication of the original matrix with its transpose, we get the (symmetrical) publication-correlation matrix

$$\begin{aligned} \mathbf{N} \cdot \mathbf{N}^{\mathrm{T}} &= \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{5} & 3 & 2 & 0 \\ 3 & \mathbf{3} & 2 & 0 \\ \mathbf{2} & 2 & \mathbf{2} & 0 \\ 0 & 0 & 0 & \mathbf{0} \end{pmatrix} = \mathbf{P} \end{aligned} \tag{10.3}$$

Now we have transferred the PAN (represented by matrix $\mathbf{N}$) into the PN (represented by matrix $\mathbf{P}$). The *diagonal* values (again in boldface) of this matrix $\mathbf{P}$, $p(i,i)$ indicate the number of attributes in each publication. For p_1 this is $p(1,1) = 5$, for p_2 we find $p(2,2) = 3$, and so on. The *off-diagonal* values give the number of attributes shared by any two publications; for instance, p_1 and p_3 (value in red) share $p(1,3) = 2$ attributes (namely a_1 and a_4, see Fig. 10.1). Thus, $\mathbf{P}$ provides the *strengths* of the links between each possible publication pair on the basis of how many attributes a publication pair has in common. In the case the attributes are cited papers, the off-diagonal values give the (not-normalized) bibliographic coupling strengths.

To keep our explanation simple, we do not go on with further mathematical elaboration, for instance the calculation of similarities by size-normalization of the strengths and other methods of normalization of co-occurrence matrices. The discussion on the best normalization procedure is endless and goes on for over 35 years; we refer to the relevant literature, for instance [10.104, 105].

If the attributes are concepts—for instance, keywords given by the author(s) and/or by the database, or terms and noun phrases identified by grammatical parsing of a publication's title and abstract—the AN represents a co-word map. We will discuss this type of mapping in detail in Sect. 10.4. For early pioneering work on the construction of co-word maps using combined multidimensional scaling and clustering, we refer to [10.106, 107].

In this section we continue with *cited papers as attributes*, and with that we focus on the basics of citation analysis. Citation analysis can be defined as the collection of all thinkable measures enabled by the PAN in which the attributes are cited papers. It covers the broad spectrum from what is often seen as simple counting to sophisticated citation-based mapping. As one might have guessed, the assumed simple counting is not that simple. It can be done in many ways: for example, the number of citations for a single publication or for an author, group, program, institute, university, country, field; in a specific period of time; from specific fields and institutions; the total distribution of citations over publications in a field. The idea (not infrequently of clients of bibliometric work) that citation analysis is just simple counting is unfortunate; it is one of the main sources of bad practice in the application of bibliometric methods for research performance assessment.

Citation-based mapping is currently in a phase of rapid development, with new application opportunities, particularly the building of a publication-level classification system of science. Analysis of the PAN—which represents direct citation (DC)—applied to the entire WoS database in the period from 2000 to the present has recently been used to construct a large-scale clustering of publications [10.108–111], representing nearly 20 million publications with around 300 million citation relations in the period 2000–2015. About 8% of the publications are excluded due to insufficient citation links. With new, advanced mathematical techniques [10.108, 109], three levels of clustering have been identified:

1. *Macro*-level clusters representing 27 broad disciplines.
2. *Meso*-level with 817 major fields.
3. *Micro*-level with 4113 fields. The micro-level fields are used for citation-impact normalization in, for instance, the Leiden Ranking [10.66, 67].

Each cluster at the micro level can be interpreted as a research field in the scientific literature. Publications are uniquely assigned to one field. The 4113 fields cover all scientific disciplines, including the social sciences. Some topics are highly interdisciplinary and encompass publications from many different research disciplines. Each of the clusters is labeled in an algorithmic way by extracting the five most relevant terms from the titles and abstracts of the publications belonging to a cluster. In this way a fine-structured citation-based taxonomy of science is created. Given its fine granularity, DC-based publication-level clustering is well suited for identifying emerging research themes by searching for new clusters and monitoring their evolution over a period of a few years after their appearance. With proper data collection and calculation algorithms, this large-scale publication-level clustering can be automated. Thus, new bibliometric tools can be constructed, for instance the *CitNetExplorer* for analyzing and visualizing citation networks [10.110, 112] and the *VOSviewer* for creating science maps based on citation or concept relations [10.113, 114]. We will come back to these bibliometric tools and other applications of publication-level clustering for science mapping in Sect. 10.4.

In Fig. 10.2, an example is shown of a small part of the entire science network at the lowest clustering level: field 175, scientometrics/bibliometrics [10.108, 109]. It covers 12 717 publications; in the figure, the 5910 publications with more than 10 citation relations are represented. The titles of these publications are given (first 30 characters), with the size of the letters proportional to the number of citations. The network is created with the VOSviewer. Each cluster within this field is indicated by a different color; the creation of clusters and their colors follows automatically from the citation relations in the network. We clearly see that the clusters within the field represent different re-

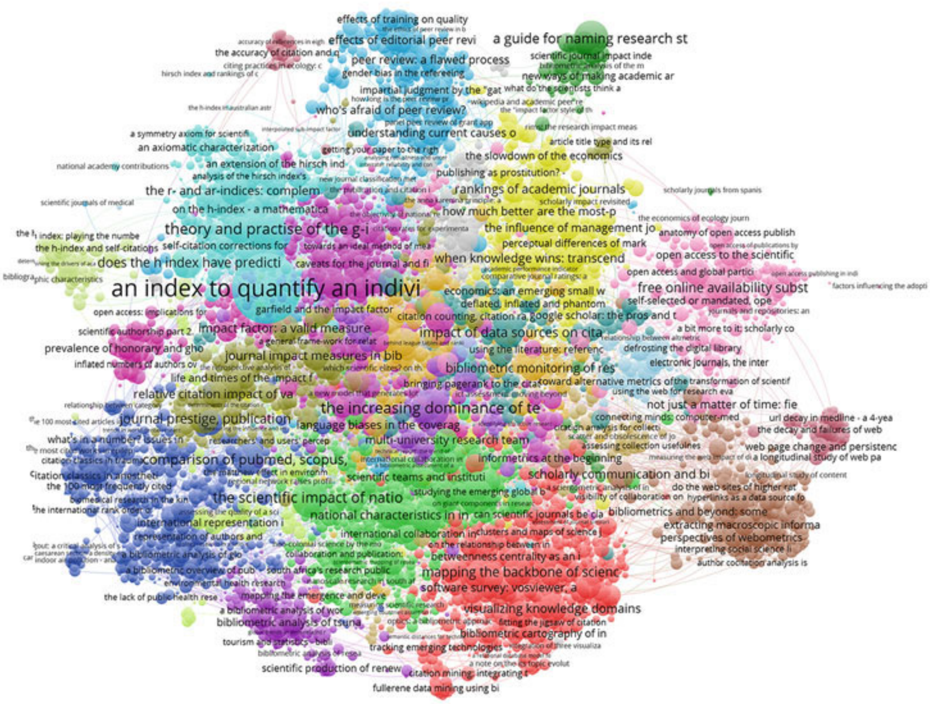


Fig. 10.2 Map of the publication-level network showing the field of scientometrics/bibliometrics

search themes within scientometrics/bibliometrics. For instance, the red cluster in the lowest part of the figure relates to bibliometric mapping; the blue cluster in the upper part relates to peer review; the light purple cluster on the right side relates to open access; the light blue cluster on the left side relates to the h-index. We see two prominent papers: *An index to quantify an individual's scientific research output*, the paper in which *Jorge Hirsch* launched his h-index [10.72] (cited 3474 times in the WoS Core Collection through December 1, 2018); and *Theory and practise of the g-index* of *Leo Egghe* [10.115] (cited 770 times in the WoS Core Collection through December 1, 2018). We stress that this publication-level clustering is *journal-independent*: all relevant scientometrics/bibliometrics publications are covered, including those in journals other than the typical scientometric journals. The above-mentioned highly cited Hirsch paper, published in the *Proceedings of the National Academy of Sciences of the United States of America*, is a striking example.

On the basis of a comparison of the three standard citation-based analyses (DC, CC, or BC) it was shown that DC provides a more accurate taxonomic representation of science than either BC or AC [10.116]. This does not imply, however, that the accuracy of DC-based publication-level classification cannot be enhanced by additional BC-based local clustering. Comparisons of different methods for mapping scientific publications on the basis of citation relations focus specifically on clustering methods, community detection, and hierarchical map-equation methods [10.117–119].

10.2.4 Indicators of Research Output and Impact

The Construction of Indicators

The central task in bibliometric research is the development of methods and techniques for the design, construction, and application of quantitative indicators on important aspects of science. What is the difference between *data* and *indicators*? How do we convert data into indicators? An indicator is the result of a specific mathematical operation (often straightforward arithmetic) with data. For the very basic indicators, there is hardly any difference from data, for instance the mere number of publications or citations of a research group. Clearly, the difference between data and indicators becomes more pronounced if citation counts of all publications of a research group in a particular field are normalized to citation counts of all publications worldwide in the same field. An indicator is a measure that explicitly addresses some assumptions. In our example, the assumption is that this is the way to calculate the international scientific influence of a research group. This raises an interesting question: what features of science can be given a numerical expression? Thus, indicators cannot exist without a specific goal in mind. They must be problem-driven; otherwise they are useless [10.1]. They have to address specific questions such as those presented in Sect. 10.2. In this way, indicators enable us to gauge important driving forces in science—for example, how scientific progress is related to specific cognitive (e. g., open questions in science) as well as socioeconomic aspects (e. g., opportunities to apply research results). Indicators have to describe the recent past in such a way that they can inform us about the near future.

The above suggests a fundamental role of indicators: their ability to test aspects of theories and models of scientific development and its interaction with society. In this sense, indicators are not only performance assessment and mapping tools for science policymakers and research managers, but also instruments in the study of science. *Price* [10.120] strikingly described the mission of the indicator-maker: find the simplest pattern in the data at hand, and then look for the more complex patterns that modify the first. What should be constructed from the data is not a number but a pattern, a cluster of points on a map, a peak on a graph, a correlation of significant elements on a matrix, a qualitative similarity between two histograms. If these patterns are found, the next step is to suggest models that produce such patterns and to test these models with further data. A numerical indicator or an indicative pattern, standing alone, has little significance. The indicators must give perspective: the change in an indicator with time, different rates of change in two different indicators, and numerical quantities replaced by geometrical or topological objects or relations [10.121]. In principle, bibliometric indicators can be seen as vectors in a three-dimensional space. The first dimension relates to performers: individuals, research groups, universities, countries, or combinations of these such as all universities in the EU countries. The second dimension relates to aspects: output, impact, collaboration, or combinations of these such as output and impact in international collaboration. The third dimension relates to subjects: research themes, journals, fields, broader disciplines, the whole of science. This performer-aspect-subject space is helpful in understanding relations between bibliometric indicators. And of course, as in the real world, the fourth dimension relates to time.

We start simply but will be confronted soon with the complications associated with creating indicators. The basic indicators, number of publications and citations, are illustrated by Fig. 10.3. At first sight it might appear to be just counting numbers. But the

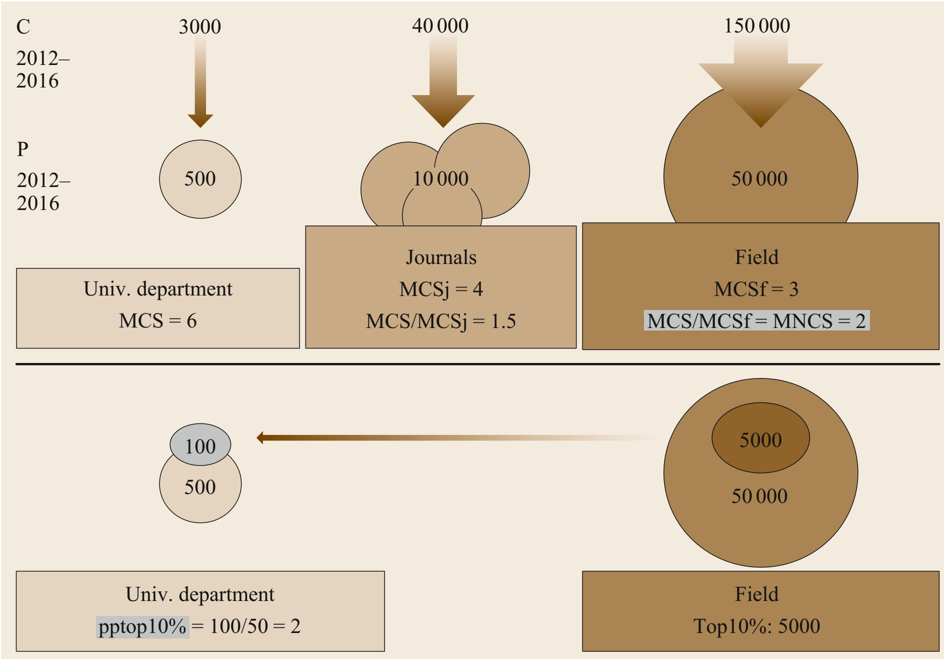


Fig. 10.3 Basic and normalized bibliometric indicators. The *upper part* of the figure shows the indicators based on averages, the *lower part* indicators based on the citation distribution function. See Schemes 10.3 and 10.4 for further explanation

reliable determination of even these two most basic indicators is far from trivial. Verification is crucial in order to remove errors and to detect incompleteness of addresses of universities, departments, and groups, and to ensure correct assignment of publications to research groups and completeness of publication sets. *Standardized procedures* for carrying out the analysis as conscientiously as possible are absolutely crucial. These procedures must be thoroughly discussed beforehand with the institutes concerned. This approach has been a long-standing common practice at CWTS [10.1, 122, 123]. For its data analysis and the calculation of bibliometric indicators, CWTS uses a dedicated bibliometric database which is an improved and enriched SQL-based version of the WoS database (*SQL*, Structured Query Language, is a standardized language used in programming and designed for managing data stored in a relational database management system). It is important to stress that a bibliometric database must have an open, highly interoperable structure [10.124].

Most readers, I am sure, are familiar with the two types of indicators that can be distinguished: indicators based on *averages* and indicators based on the *citation distribution* within fields. The latter approach is important, because the distribution of citations over publications is skewed [10.54, 125]. The indicators are explained in more detail with an example in Schemes 10.3 and 10.4. The normalization procedure is particularly crucial [10.126]. Moreover, research groups publish in more than one journal, and they are active in more than one field. Therefore, weighted average values have to be calculated, the weights being determined by the total number of publications published by the institute or department in each journal or field.

Scheme 10.3 Indicators Based on Averages

We showed in Fig. 10.3 a university department (or institute, research program, university) that has 500 publications in the recent 5-year period 2012–2016 ($P = 500$). This is the research output defined as the number of articles of the department in the given period as far as coverage by the WoS (or Scopus). We consider as articles the following publication types: normal articles (including proceedings papers published in journals), letters, notes, and reviews (but not meeting abstracts, obituaries, corrections, editorials, etc.). Within the same period, these 500 publications are cited 3000 times ($C = 3000$). This is the total impact defined as the number of publications that cite in the given period the research output of the department.

The average number of citations per publication MCS (mean citation score) $= 6$. This is the size-normalized citation impact of the department. It is important to be clear as to how publications and citations are counted: the counting procedure is directly related to the definition of the *citation window*. We discuss the counting procedure in Schemes 10.5. For field-specific normalization purposes, the citation impact of a department is compared with similar measures for:

1. *All journals used for publication by the department* (in the example in Fig. 10.3, these journals cover in total 10 000 publications that are cited 40 000 times): the journal impact MCSj (mean citation score of the journals based on a weighted average, measured over the same 5-year period and taking article type into account) in this example MCSj = 4.
2. *All journals in all fields in which the department is active* (in total 50 000 publications that are cited 150 000 times): the field average MCSf (mean citation score of all journals in all fields based on a weighted average, again measured over the same 5-year period and taking article type into account). In this example, MCSf = 3.

We observe the following. The department performs better than both the journal and the field average, MCS/MCSj = 1.5 and MCS/MCSf = 2.0, respectively. The latter indicator, the average number of citations of the publications normalized for field differences and publication year, is for the sake of brevity written as MNCS. We also see that the journals chosen by the department for publications are the journals with an above-average impact in the fields: MCSj/MCSf = 1.3. We call this indicator the mean normalized journal score, for the sake of brevity written as MNJS.

The above indicators are a simple representation of the normalization procedure. In reality, this normalization procedure is somewhat more complicated [10.127–129]. We confine ourselves to remarking that normalization of the number of citations is in principle a comparison of the *real number* of citations with an *expected number* of citations based on, for instance, the average number of citation of all publications published in the same journal or field, in the same year, and with the same document type.

For many years, the MNCS and its predecessor [10.127–129] have commonly been regarded as the *crown indicator*, because this indicator directly measures the extent to which a research group, department, institute, research program, or university performs significantly above the international level of their own field(s). From long-standing experience, we know that at the department level, an MNCS value above 2.0 indicates a very strong department, and above 3.0 the department can be generally considered to be excellent and comparable to the top departments in the same field at the best US universities. Thus, if the threshold value for the MNCS indicator is set at 3.0, excellent groups can be identified with high probability. Because the above indicators, and particularly the MNCS, are based on averages, they are sensitive to outliers. This less attractive property may lead to the *Göttingen effect* discussed in Sect. 10.3.

Scheme 10.4 Indicators Based on the Citation Distribution within Fields

Again we look at Fig. 10.3, *lower part*. The indicator ppttop10% is the proportion of the publications of a department (or institute, program, university) that, compared with other publications in the same field and in the same year, belong to the top 10% most frequently cited. In the example, the field has 50 000 publications in the given time period (2012–2016), of which by definition 5000 belong to the top 10% of the citation distribution. Say 100 of these 5000 are publications of the department which has a total of 500 publications. Thus, the statistical *expectation value* for the number of top-10% publications is 50, but the department has twice that number, so the pptop10% for this department is 2.

Given the earlier-mentioned skewed distribution of citations over publications, we regard the pptop10% indicator as the most appropriate bibliometric impact measure. Generally, the correlation between the average-based MNCS and the distribution-based pptop10% is high, particularly at higher aggregation levels such as larger research institutes and universities [10.66, Fig.2]. This distribution-based indicator is insensitive to outliers (we again refer to the discussion of the Göttingen effect). With a statistical technique known as bootstrapping, 95%-stability intervals can be calculated. A stability interval indicates a range of values of an indicator that are likely to be observed when the underlying set of publications changes to a specific extent. For an example of the stability intervals of the pptop10% indicator, we refer to the methodology section of the Leiden Ranking [10.67]. Of course, the top 10% is a rather arbitrary choice. But, at least at the level of universities, pptop5% or pptop20% give very similar results as pptop10%. Thus the top threshold is not that important in this range; for empirical results we refer to [10.66].

Scientific fields differ substantially in citation practices: differences in average number of references, in average percentage of references to recent publications and of references to other fields, in the coverage of publications by a database such as WoS or Scopus, and in the growth rate of a field. These differences can be very large. For instance, the average number of references (i. e., citations given) in biochemistry & molecular biology is about an order of magnitude larger than that in mathematics. In other words, there are large differences in citation density between fields. But even *within* fields, citation density may vary considerably [10.130], as we will see in Sect. 10.4. Therefore, inevitably, in normalization procedures for citation-based indicators, such differences have to be taken into account for meaningful comparisons of citation impact between fields and within a field. In our example, we discussed how the MNCS indicator and the pptop10% indicator correct for differences in citation practices between scientific fields based on the WoS field journal-based classification system. As illustrated in Fig. 10.3, each publication belongs to one or more fields, and the citation impact of a publication is calculated relative to the other publications in the same field(s). The accuracy of journal-based classification of fields is an issue of continuous discussion [10.131], particularly the extent to which the classification system allows for overlap of categories. We note that indicators such as MNCS and pptop% will have different values for different field definitions (for instance, the journal-based WoS categories versus the publication-based micro-level fields), because the calculated normalization depends on the field definition.

Because the pptop10% indicator is not sensitive to outliers, it is more stable than the MNCS indicator. In that sense, the pptop10% indicator has taken over the role of crown indicator. With an advanced bibliometric data system and proper algorithms, it is no problem to identify for instance the top-1% publications. Often the idea is that these very highly cited papers mark new developments, scientific breakthroughs. But be careful. A top-1% publication is not necessarily a breakthrough or highly innovative paper. Also, in science, there are hypes and trends, and a very highly cited paper could be a paper that fits well in a specific trend which may be short-lasting. Furthermore, papers in which extended reviews or important research methods and algorithms are presented can be very highly cited. Such papers, notwithstanding their importance, are more mainstream and not necessarily innovative.

In another, recently developed approach called *source normalization* (SNIP: source normalized impact per paper, also called *citing side normalization*), the normalization procedure is applied by analyzing the referencing behavior of citing publications. Thus a specific publication, or set of publications, is, as it were, compared with its own environment, namely the references of all papers citing the specific publications, instead of comparison with the publications across the entire field. In this way, SNIP corrects for differences in citation practices between scientific fields [10.132–134]. An extensive discussion of the SNIP approach and its variants is given in [10.135].

A recent review of the literature on citation-impact indicators is given in [10.136]. The above-discussed impact indicators have been regularly criticized, particularly in terms of the relation between these indicators and input, i. e., resources such as number of staff, experience of staff, or available funding. The opposition to the MNCS indicator, particularly its size-independent character, recently took the form of a crusade [10.137, 138], and this provoked many counterreactions [10.139–146]. In main lines, the outcome of this debate is that input normalization is very tricky, because the assessment of, for instance, time spent on research cannot be measured accurately and is easy to manipulate. Furthermore, in practice, the MNCS works quite well; its calculation is relatively simple, transparent and straightforward. This is an important user value, and in this respect there are really no better alternatives. Moreover, in any research evaluation procedure, not just one but a set of indicators must be used (including the pptop10%, see our real-life example in Sect. 10.3). This provides a broader view of the performance of a research group or institute under evaluation [10.140]. Curiously, this debate ignores the extensive work regarding the effects of size-independent indicators on size-dependent indicators [10.147, 148].

Counting and Weighting Procedures

We already remarked that nothing seems to be simpler than counting. But the precise way in which publications and citations are counted is a crucial part of the construction of bibliometric indicators. In Scheme 10.5, we present a practical approach to publication and citation counting, taking citation window length and evaluation period length into account.

Scheme 10.5 Example of a practical approach to publication and citation counting taking citation window length and evaluation period length into account

There is ample empirical evidence that in the natural and life sciences, the peak in the number of citations occurs on average in the third or fourth year after publication. A 4-year period is therefore appropriate as a *citation window* for impact assessment. If longer-term impact is expected to be important—which is often the case for engineering, social sciences, and humanities—the time period of the citation window and the time blocks used in the trend analysis need to be broadened. Note that the length of the citation window is not the same as the length of a total evaluation period (which can be short or long).

How can the length of the citation window be combined with the total length of the evaluation period while including the most recent citation year?

For a trend analysis, the *roof-tile approach* is often used: rolling and partially overlapping 4-year periods (blocks) in order to smooth out insignificant annual fluctuations. For instance, for the 10-year evaluation period 2007–2016, 2007–2010 is the first 4-year block (roof tile), followed by 2008–2011, and the last block is 2013–2016. The counting procedure is as follows. We start with the first block 2007–2010. For the 2007 papers, citations are counted during 2007–2010, for 2008 papers citations during 2008–2010, for 2009 papers citations during 2009–2010, and for 2010 papers the citations only in 2010. Then, in the next block everything shifts 1 year, so now for the 2010 papers citations will be counted in 2010 and 2011. Finally, in the last block 2013–2016, citations to the 2013 papers will be counted during 2013–2016, for the 2014 papers during 2014–2016, for the 2015 papers during 2015–2016, and for the 2016 papers during 2016.

We also note that the importance of a publication does not necessarily appear immediately, even to peers, and that specification of quality may take considerable time. Therefore, it is important to make assessments with shorter as well as longer time windows. Extreme and relatively rare cases of delayed recognition are sleeping beauties, mentioned earlier (see also Sect. 10.3).

As we discussed above, a clearly described publication- and citation-counting procedure is a crucial and absolutely necessary technical element in bibliometric analysis. Is this the end of our counting story? Certainly not. We are now faced with the more fundamental question of whether publications and citations should be counted with integer 1. This is the problem of *full* versus *fractional* counting, which is in fact a form of weighting. Indicators can be calculated using either a full counting method or a fractional counting method. The full counting method gives equal weight (with value 1) to all publications of a specific entity such as a group, department, institute, or university, regardless of collaboration. The same goes for the citations received by these publications. The fractional counting method, however, gives less weight to collaborative publications than to non-collaborative ones, simply because the publications (and their citations) are divided over the collaborating institutes. The counting modality influences normalization; for discussions on the relation between fractional counting and field normalization, we refer to [10.133, 149, 150]. The fractional counting method leads to a more proper field normalization of impact indicators [10.151, 152]. We emphasize that in an advanced bibliometric system equipped with the proper algorithms, it is not a problem to calculate indicators with both full and factional counting. Because of the better normalization properties, fractional counting is regarded as the preferred method in the Leiden Ranking, but both modalities are available in this ranking [10.66]. The advantage in having both is that it provides a good idea of the robustness of the outcomes. At higher aggregation levels such as universities, the correlation between the ranking based on full counting and that based on fractional counting is high [10.66].

Indicators of Research Collaboration

It is normal bibliometric practice to calculate a set of indicators of *scientific collaboration* [10.66, 67]. The most general one is P(collab) and pp(collab), the number and proportion, respectively, of inter-institutional collaborative publications—for example, the proportion of the publications of a university that have been co-authored with one or more other organizations. But collaboration has many aspects. For instance, we can distinguish collaboration at a national or international level, between universities and the business sector, and collaboration at smaller or larger distance. Therefore, next to the above general indicator, we can extend the set of collaboration indicators with a number of specific collaboration indicators. First, international collaboration: P(intcollab) and pp(intcollab) are the number and proportion of international collaborative publications—for example, the proportion of the publications of a university that have been co-authored by two or more countries. Second, collaboration with the business sector: P(UIcollab) and pp(UIcollab) give the number and proportion of collaborative publications with industry, such as the proportion of the publications of a university or within a research program that have been co-authored with one or more industrial partners.

Although it has been demonstrated many times that international collaborations generally produce more highly cited science than national collaborations, the physical distance between collaborating research groups remains an interesting variable. Recent work showed that physical proximity of collaborators positively influences the scientific impact of their publications [10.153, 154]. As a first step in taking collaboration distance into account in performance studies, distance indicators can be calculated based on the largest geographical distance between two addresses mentioned in each of the publications—for instance, P(<100 km) and pp(<100 km), the number and proportion of relatively short-distance collaborative publications; and P(>5000 km) and pp(>5000 km), the number and proportion of long-distance collaborative publications. Using the Leiden Ranking menu [10.66], one immediately sees that much of the Australian international collaboration involves long distances. Collaboration indicators are always calculated using the full counting method [10.67]. There is a vast amount of literature on the many aspects of scientific collaboration. An overview is given in [10.155]; for recent work on university-industry collaboration, we refer to [10.156].

Indicators of the Growth of Science

We conclude Sect. 10.2 with a small excursion outside the world of research performance. Measuring science also entails the *measurement of properties of science as a whole* or of its main fields. Examples are scientific development in terms of growth, aging of literature, statistical properties of important distribution functions, size distribution of clustered publications or journals, and so on. Section 10.4 of this chapter discusses the mapping of science and therefore deals with several such macro issues. Here we confine ourselves to the first two topics mentioned: growth of science and aging of literature. Researchers prefer to cite the more recent work. Thus, analyzing the number of references as a function of their age, i. e., as a function of the publication year of the cited publication, enables us to measure the aging of scientific literature and the differences in aging between the disciplines. However, in earlier times, there were much fewer documents published than in recent times. Thus, the time-dependent distribution of references will always be a specific combination of aging and growth of the scientific literature. How to disentangle these two independent phenomena? Empirical work shows that a typical aging process is particularly active in an exponential fashion for a period of up to about 10 years back from the publication year of the citing publications. For earlier years, it appears that this aging process is much less strong, so that either aging is still present but with a much slower exponent, or, in a first approximation, there is no longer any aging. In the latter case, the number of references will be approximately proportional to the number of papers available at that time, and provides a measure for the growth of the science [10.157].

10.3 Practical Application of Research Performance Indicators

It is also not unreasonable to expect that in citation-based performance analysis, pitfalls and sources of error will always be with us. In the *Leiden Manifesto* [10.158], principles to guide the use of research metrics, in particular bibliometrics, are formulated. Most of the guidelines were already discussed some 20 years ago [10.159, 160], but the recent increase in a rather uncontrolled use of bibliometric indicators forced the bibliometric community to formulate, loud and clear, the principles of proper use of bibliometric methods. I think we have to see whether this will work. Very recently, an experienced member of EU evaluation panels remarked

> You cannot imagine the extent of less good practices, which, however, seem to have become very popular everywhere (such as the use of h-index or impact factors in connection with individual researchers).

10.3.1 Methodological and Technical Issues in Evaluation Studies

Methodological and technical issues must first be recognized, after which, where possible, an adequate solution to the problem must be found. Here we present a list of ten important issues that must be taken into account in every bibliometric analysis:

1. *Applicability of bibliometric methods* in the different scientific disciplines, particularly in engineering, social sciences, and humanities. This applicability is directly related to the coverage of publications by the WoS or Scopus database (coverage in relation to the list of publications provided, for instance, in annual research reports). If this coverage is too low—for example, less than 50%—bibliometric analysis based on citations will only give a partial picture of the research performance [10.91–96]. Another increasingly important problem is the need to correct the bibliometric data for *retracted papers* [10.161–163].
2. *Influence of language*, particularly German and French, on the assessment of research performance and on the ranking of universities [10.164, 165]. Non-English publications count as output if they are covered by WoS or Scopus, but their impact is usually very low. This causes considerable dilution of the average impact. In order to avoid this problem, one has to conduct the analysis with English publications only. See also the discussion on core publications in [10.67].
3. *Length of citation windows*: important publications may be cited after many years. This is the phenomenon of delayed recognition [10.98, 99] or, in extreme cases, *sleeping beauties* [10.166–177]. Generally, evaluation periods are restricted to rel-

atively short periods in order to assess research performance as recent as possible, for instance the last 5 years. Although the probability is not high, it is possible that scientific work finds recognition long after publication, for instance after 8 years or more. Therefore, where possible, one should also use longer citation windows.

4. *Statistical properties* of bibliometric indicators, for example, their skewness and scaling behavior [10.147, 148, 178]. We already discussed the importance of taking into account the skewed distribution of citations over publications for the construction of indicators. The scientific productivity of researchers will also often show a skewed distribution in larger institutions. This is relevant when using input indicators based on averages.
5. *Concerns about the journal impact factor* [10.60–62, 179] *and h-index* [10.72, 180–182] for accurate research performance assessment. These easily available, low-cost indicators are often used to perform a quick, simple, and cheap bibliometric research evaluation. They may give first impressions of performance, but cannot replace advanced bibliometric methods.
6. *Relation between peer review* judgment and bibliometric findings [10.41, 147, 180, 183–186]. This is an important issue in research performance assessment. Neither peer review nor bibliometric analysis is free from shortcomings. Furthermore, bibliometric findings and outcomes of peer review are not independent variables in the quality judgment space: peers use bibliometric elements in their judgment; for instance, they generally attach great value to publications in top journals.
7. *Open access*, *webometrics* and *altmetrics* [10.187–205]. The ever-increasing accessibility and availability of data on research activities has created new forms of research metrics, which quickly become important for the analysis of different aspects of research performance. But this type of metrics is also not problem-free.
8. *Definition and delimitation* of fields as well as *assignment of publications* to these fields or to other specific entities such as research groups, institutes, or universities. The problem of accurate assignment of publications is often underestimated. It is, for instance, far from trivial to correctly collect the publications of a university [10.67].
9. *Large differences in citation density between and even within major fields* demand the use of advanced normalization procedures [10.130]. We again stress the crucial importance of proper normalization in the construction of indicators, as we did in the preceding section [10.126]. It is important to realize that field definitions determine the outcomes of the normalization procedure and, with that, the values of indicators.
10. *Rankings of universities*, though controversial, enjoy a great popularity. They are probably influential in guiding the choice of young scientists and research students [10.63–71]. For interpretation of ranking results, it is crucial to know all relevant properties and, particularly, the methodological and technical questions of these ranking properties.

In the next sections we discuss the following issues in greater detail: concerns about journal impact factors and h-index for research performance assessment; assignment of publications to research groups, institutes, and universities; relation between peer review judgment and bibliometric findings; role of open access, webometrics, and altmetrics; and ranking of universities. We conclude this section with a discussion on arguments against bibliometric analysis in general.

Journal Impact Factor

The definition of the *journal impact factor* (JIF) can be described as follows. We take as an example the year 2016. The JIF of 2016 for a specific journal is the total number of citations (given by articles in all possible journals covered in the WoS) received in 2016 by the publications of 2014 and 2015 in the journal, divided by the total number of these publications of 2014 and 2015.

Simplistic or *easy-to-go* approaches in the use of bibliometric data are popular in many research evaluation practices, particularly the use of the abovementioned JIF. In this *poor man's citation analysis*, publications of a research group are weighted with the impact factors of the journals in which the publications have appeared. In this way, a total score is calculated and thus a ranking of groups can be established. This ranking can then be used as a basis for financing. Bibliometric researchers have for more than 20 years explicitly warned against the use of journal impact factors for research evaluation purposes [10.60–62]. Weighting publications with impact factors is usually a poor estimate of the actual impact of (a group of) publications, because of (1) the *skewed distribution of citations* over individual publications in a journal, and (2) the fact that impact factors are strongly influenced by *review papers*, which are often more highly cited than normal publications. This latter point is also a general issue: in normalization procedures, article type has to be taken into account. The use of journal impact factors has an equalizing effect: two groups A and B publishing in the same journals have the same impact, as their publications are weighted with the same impact factors, although the publications of group A are much more frequently cited than those of group B. Moreover, journal impact factors can be heavily affected and even manipulated by journal self-citation [10.206].

We note that the skewed distribution of citations over individual publications also applies not only in journals, but for any set of publications—for instance, in research groups. This is the reason that citation-distribution-based impact indicators such as the pp-top10% are to be preferred over the citation-average-based indicators such as the MNCS. On the other hand, already at a quite low aggregation level, such as a research group, the correlation between the average citation score of the publications of the group (actual impact) and the average citation score of all articles in all journals used by the research group (journal-based proxy of impact) is high [10.147]. The concerns about the JIF are related to the use of this indicator as a measure of the impact of individual publications within a journal. In spite of the availability of reliable bibliometric methods, journal impact factors are still commonly used in research evaluation practices. Fortunately, researchers and journal editors are becoming increasingly aware of this problem, as evidenced in the *San Francisco Declaration on Research Assessment* (DORA) [10.207]. The bibliometric community made their position forcefully clear in the earlier-mentioned *Leiden Manifesto* [10.158].

Journal impact indicators are certainly not useless. They should be used for the goal for which they were developed: a measure of journal prestige. Nevertheless, recent research [10.208] takes a provocative stance and shows that the use of the JIF for assessing individual articles need not necessarily be wrong. Other indicators of journal performance have been developed, including the *cited-side-normalized* indicator MNJS discussed in Sect. 10.2. This journal indicator also correlates significantly at a higher level—e. g., all publications of a research group—with the actual impact, i. e., the average citation score of the publications of the group (see [10.147]; here, in fact, the predecessor of the MNJS indicator was used). Furthermore, as discussed earlier, *citing-side-normalized* indicators have been developed, such as the SNIP indicator and the audience indicator [10.209, 210]. Recently, CWTS developed a free-access SNIP journal indicator application tool based on Scopus data for about 20 000 journals [10.211]. Scimago developed its Scopus-based journal indicator [10.212, 213], and Elsevier launched the Scopus-based Elsevier CiteScore, which is quite similar to the JIF, but it takes all document types into account in a more consistent manner [10.214]. Data quality determines the quality of the indicators. Recently, authors have discussed a typology of errors and inaccuracies in the Scopus database [10.215, 216].

Shortly after the launch of the JIF [10.22], a recursive journal performance indicator, the influence weight indicator, was introduced [10.24]. Twenty years later, this idea returned in the form of the PageRank algorithm for the Google search engine [10.217]. This inspired the construction of new recursive indicators such as the Eigenfactor [10.218]. For a comparison of these indicators, we refer to [10.210].

The h-index

A remarkable event in the history of bibliometrics is the introduction of the *Hirsch-* or *h-index* named after its creator *Jorge Hirsch* [10.72]. This new indicator attracted enormous attention, particularly for its charmingly simple structure: an author has an h-index with value n if n of his/her papers have at least n citations each and the other papers have at most n citations each. A torrent of publications describing numerous variants of the h-index followed [10.115, 181]. The attractiveness of this measure lies in the ease with which it can be determined: by simply ranking in the WoS or Scopus the publications of a researcher by the number of citations received, one will find a publication with ranking number equal to the number of citations it received. For instance, for the author of this chapter, his publication with ranking number 42 in the WoS has 44 citations, whereas the publication with ranking number 43 has 42 citations, thus fewer citations than the ranking number. It follows that the h-index of the author is 42 (December 1, 2018).

Although the h-index gives a first impression of a researcher's impact, we advise against its use as a basis for comparison of researchers, for the following reasons. First, the h-index is a typical *lifetime indicator*, strongly dependent on the total number of a researcher's publications. This means that the h-index will most probably put younger researchers at a disadvantage. Second, as we already noted, there are large differences in citation density between and even within research fields. In medical fields, the clinical subfields often have significantly lower citation density than the basic science subfields [10.130]. Let us give a practical example. For a new chair in neurology, a university must choose between two candidates. One is more clinically oriented, the other more basic. If the choice is based on the h-index, the more clinically oriented candidate will be at a disadvantage, because his/her h-index will most likely be lower than the h-index of the other candidate, given the large difference in citation density between the clinical and basic fields of neurology. Third, *Waltman* and *van Eck* have shown that the h-index is mathematically inconsistent, and illustrate this with a simple example [10.182]. Fourth, a simple determination of the h-index as described above will include self-citations and is therefore open to manipulation. For a comparison of the h-index with bibliometric indicators as discussed in Sect. 10.2 in the performance assessment of a large set of research groups, we refer to [10.180]. An interesting experiment is the development of agent-based models to describe the process of generating publications and citations including a prediction of the h-index [10.219].

Assignment of Publications

In evaluation and monitoring procedures, it is crucially important to know which results—for instance, publications—have to be assigned to a specific research entity (group, institute, university, program, etc.). This is not always that easy. The problem is largely related to a proper definition and delimitation of the research entity. We take universities as an example. Even at this high aggregation level, the assignment of publications is not a straightforward task. A university may be referred to using many different (often non-English) name variants and abbreviations. In addition, the definition and delimitation of universities as separate entities is not always obvious. International differences in the organization of academic systems also poses difficulties in terms of identifying the proper unit of analysis. In the following, we draw our discussion largely from the text in the methodology section of the *Leiden Ranking 2016* [10.67].

Instead of applying formal criteria (if possible at all), it is more effective to follow common practice based on the way the institutions are perceived locally. For instance, the University of Cambridge is to be considered as one entity, but in the case of the University of London, one must distinguish between the constituent colleges. For the United States, university systems (e. g., University of California) have been split into separate autonomous universities. The higher education sector in France, as in many other countries, has gone through many reorganizations in recent years. Many French institutions of higher learning have been grouped together. However, for comparison of university performance, one must distinguish between the different constituent institutions, except in the case of full mergers. Furthermore, universities may be split into various entities, for instance the Austrian universities where medical faculties became separate universities in 2004; or the reverse, such as the University of Lisbon merged in 2013 with the Technical University of Lisbon.

A major problem relates to affiliated institutions, i. e., research institutes and, in particular, hospitals associated with universities. Among academic systems, wide variation exists in the types of relations maintained by universities with these affiliated institutions. Typically, these relationships are shaped by local regulations and practices, and affect the comparability of universities on a global scale. As there is no easy solution to this issue, it is important that a transparent methodology is used in the treatment of affiliated institutions. Particularly in the case of university hospitals, the correct or incorrect assignment of publications to a university may have significant consequences, given the often large output of medical research. Specific name-based assignment algorithms are necessary, because researchers employed by a university but working at affiliated institutions may not always mention their university in publications. Universities have become increasingly aware of this problem, which affects their visibility in research publications, and they actively exert pressure on researchers to mention their affiliation with the university in publications. CWTS developed a practical approach in which three different types of affiliated institutions are distinguished; we refer to [10.67] for details.

Peer Review and Bibliometric Findings

One of the first questions that comes to mind when speaking about research evaluation is, how do bibliometric findings relate to peer review judgments? On the basis of ample experience [10.185], we find that there is generally a significant correlation between the opinions of peers and the results of indicators [10.122]. This is even the case for young scientists: studies on PhD graduations show that *cum laude* graduates have a significantly higher citation impact than graduates without the honor degree [10.83, 84]. A study on peer review and bibliometrics in the field of condensed-matter physics [10.183] showed that the number of publications correlates only weakly with peer review judgments, whereas the field-normalized citation indicator shows a considerably stronger correlation. This would indicate, at least for the field studied, the absence of a *publish-or-perish* mentality. The often-heard presumption that both bibliometric analysis and peer review would undervalue interdisciplinary research finds no empirical support, at least not in physics [10.184].

Correlations are statistical measures based on large numbers. It is very well possible that peer review judgment and indicators may disagree substantially in specific cases. Based on long-standing experience, we estimate that, in the case of research groups, this happens in about 25–30% of cases. In half of these, peers have a positive judgment and the indicators show low performance, and in the other half the opposite is true. These are of course interesting cases to determine what is going on. Did the bibliometric analysis not grasp the essence of a group's performance? Are the peers well informed? Both bibliometric analysis and peer review have their shortcomings. Some authors even state that peer review suppresses scientific innovation [10.85]. Other authors have claimed that peer review reinforces nepotism and sexism, and that it puts women at a disadvantage [10.86]. Further research, however, does not support these assertions [10.87, 88]. For a study on the selection of excellent research through peer review committees and a comparison with bibliometric indicators, we refer to [10.186]. A specific problem in the comparison of peer judgments and bibliometric results is that there is scarcely any information about the dissensus within a peer committee. Thus, the statistical uncertainty of the peer judgment is largely uncharted territory.

In our opinion, the combination of peer review and bibliometric analysis is the best strategy for evaluation and monitoring procedures, at least in those research fields where bibliometric analysis can be applied. Peer review must play the central role; bibliometric analysis provides supporting instruments such as a consistent set of indicators and maps. This combination enables a reliable, objective, and transparent assessment of research performance [10.123]. Bibliometric analysis should not be used as a standalone tool. This is a decades-old statement, currently revitalized as "the humility dimension in responsible metrics" in the apologetic jargon of bibliometric circles. Nevertheless, advanced bibliometric analysis is important, because the indicators and maps provide detailed information—of which peers may not always be aware—on crucial aspects of research performance and scientific communication such as international impact, collaboration patterns, and position on the map of science. There is another good reason for using bibliometric analysis: it also stimulates hard questions that might be avoided by peers. After a pilot exercise in 2009, the Higher Education Funding Council for England (HEFCE) concluded that "citation information was insufficiently robust to be used formulaically or as a primary indicator of quality" [10.220]. No serious practitioner in advanced bibliometric analysis, however, would ever assert that this use of citation data could be the case.

The costs of an advanced bibliometric analysis are generally less that the costs of peer review procedures. Finding an appropriate team of experts in terms of reputation, experience, and size is time-consuming, and costs vary depending on the number of experts involved. Automation of data collection and data analysis reduces the costs for bibliometric analysis considerably, particularly in the case of follow-up monitoring and evaluation. Once an extensive analysis of a research group, institute, or program has been carried out, the basic data can be stored in an evaluation data system, and updating can be realized at relatively low costs. In a next phase of the evaluation procedure, the bibliometric findings and the peer judgments have to be compared, and a discussion is necessary on the differences in findings between the two approaches. In most cases, research group publications have multiple authorship. Therefore, another point of discussion, which cannot be solved bibliometrically, is the role of the individual researchers. Again, the message remains: never use bibliometric analysis, regardless of how advanced it may be, as a standalone tool for research evaluation.

Open Access, Webometrics, Altmetrics

Open access means unrestricted online access to peer-reviewed (not always), published research, primarily for journal articles, but also for a growing number of theses, book chapters, and monographs. Open access can be provided by authors by self-archiving their journal articles (not necessarily an open-access journal) in an open-access repository or in other open-access websites, for instance of their own university (*green* open access), or by publishing in an open-access journal (*gold* open access) which provides direct open access, in most cases on the publisher's website. A renowned open-access repository for physics and related fields is the preprint database *arXiv*. An example of an influential open-access journal is *PLoS ONE*. In the case of non-open-access journals, the production costs associated with publishing accepted articles lie entirely with the subscribers; in the case of open-access journals, with the authors. Open access makes scientific results more accessible, while maintaining the intellectual rights of the authors. Steven Harnad is one of the most active promoters of open access. His website offers international, up-to-date information on all possible aspects of open access [10.187]. Citation data for the Web of Science and Scopus are not free: they are provided commercially by Clarivate Analytics (until 2016 by Thomson Reuters) and Elsevier, respectively. Therefore, in a sense, the free provision of citation data by Google Scholar can also be considered a form of open access. Google Scholar offers an opportunity to use this freely accessible web search engine for the evaluation of research, particularly in the social sciences and the humanities, because it covers significantly more literature sources than simply journal articles [10.73]. Meanwhile, the WoS and Scopus have increased the coverage of books. A new methodology for comparing Google Scholar and Scopus is discussed in [10.221]. Recently, the Dimensions database was launched. This freely accessible search tool links publications to citations, grants, funding agencies, and patents [10.188, 222].

An important issue is the extent to which open access influences bibliometric properties of research publications, particularly whether open access increases the number of citations to publications [10.189]. Generally, open access is considered as a means to effective early warning, a possibility to speed up citations to publications. It is not clear yet, however, whether open access has a longer-lasting effect on citations. There are also doubts whether open access is really good for science over the longer term [10.190]. Other authors argue that open access can provide new measures of research performance based on public availability of articles, their archival location, licenses, access costs, and supporting information [10.223].

Almost two decades ago, a new way of assessing research performance emerged: *webometrics* [10.191–193]. This method (also called cybermetrics) is based on an analysis of the many connections available via the Internet, for instance the number of web connections to a university from other institutions. With a similar analysis, the Cybermetrics Lab of the Spanish National Research Council (CSIC) developed a global webometrics ranking of universities [10.194, 224]. Obviously, research institutions can exert a strong influence on the number of connections to their websites. It remains to be seen to what extent webometrics provides a reliable assessment of research performance and how it influences bibliometric measures. Related to webometrics and open access are new developments in altmetrics, short for *alternative metrics* [10.195, 196]. Altmetrics focuses on the collection and analysis of data available through the Internet [10.197], primarily data on the use of publications [10.198], views of HTML versions, and downloading of PDFs [10.199], but also mentions of publications and other forms of visibility in social media [10.200] such as scholarly blogs [10.201], Twitter [10.202], and ResearchGate [10.225].

There is fast-growing interest in altmetrics. Publishers are increasingly providing altmetrics data on papers; for example, the open-access journal group *PLoS* is active in developing standardized methods for the collection, analysis, and use of altmetrics data [10.226]. Oxford University Press provides access to the number of online views for each article, the number of citations to the article, and the Altmetric score for each article [10.227]. With Mendeley, a reader-counts database, Elsevier is strongly involved in the development of altmetrics. This reflects the commercial interest in developing alternative measurements of scientific productivity and impact. We refer to the Elsevier quarterly magazine *Research Trends* [10.203] for an overview of the major trends in altmetrics written by leading researchers in this field. Also, the websites of additional players in the field, such as Altmetrics.org [10.228] and ImpactStory [10.229], are important for keeping up with the new developments. Recent research focuses in particular on the relation between altmetrics data and citation counts [10.204]. Of course, not all of these developments are quite so harmless. Inevitably, since data from social media are also used in altmetrics, scientific and public interest are mixed. It is critical to distinguish between these two aspects, because they concern different dimensions of research [10.205]. Furthermore, it is plausible that social-media-based altmetrics is sensitive to manipulation. Like citation data, altmetrics data—and particularly the reader counts data of Mendeley—have to be normalized because of differences between fields of science [10.230, 231].

In an effort to achieve greater structure and standardization in altmetrics research, the NISO (*National Information Standards Organization*, Baltimore, USA) initiated the Alternative Assessment Metrics Project in 2013. Important issues include the development of specific definitions for altmetrics, proper usage of the term *altmetrics*, identification of best practices for aggregating multiple data sources, and the development of a statement on the role of altmetrics in research evaluation. The latest initiatives can be found on the NISO website [10.232]. Data submitted by applicants for proposals, evaluation data, data on panel members, and data on grantees can also be considered as sources of altmetrics. These data enable the combination of advanced bibliometric indicators with proposal- and grant-related data. Data extracted from the proposals could also be of importance for the realization of an effective research information system. Grantees of a research council can provide further data, for instance, acknowledgments, career development (institutional mobility, interdisciplinary mobility, leadership, training capacities), and the use of their research in socioeconomic and technological areas. In this way, a broad data system for the assessment of research can be developed with which interesting patterns in funding, mobility, and career development could be identified.

Finally we note that altmetrics-like issues have long played a role in peer-review evaluation of research groups, such as in the use of annual research or self-evaluation reports. In such reports, many forms of scientific performance are mentioned: grants from research councils and charities, contract research, lectures in international meetings, lectures for a broader public, other forms of science popularization such as radio, TV and newspaper presentations, membership of professional and social organizations and committees, and editorial work, among others.

Ranking of Universities

Scientific research has always been an international enterprise. The growth of worldwide R&D activities and of the higher education sector in developed and developing countries over the past few decades has reinforced established academic institutions and created many new ones. At the same time, the number and intensity of student and researcher exchange programs, international collaborations, and working stays outside the own country has rapidly increased, propelled by the ever-growing global mobility. An important parallel development, also stimulated by the growth in the higher education sector, is the strong demand for accountability, evidence of quality, and *value for money*. These developments have led to increased competition for funding and for the best students and researchers among universities within nations and worldwide. Universities strive to achieve top position in their country, or even the world. This implies the existence of some kind of a league (national or international) to which one can be admitted only on the basis of performance. The higher the performance, the better chances a university has to become a member of an elite league and to reach a high-ranking position in this league, which reinforces the university's appeal for the talented people.

Several organizations produce annual rankings of universities on the basis of survey data, bibliometric data, or both. We already mentioned in Sect. 10.1 the Academic Ranking of World Universities (ARWU, Shanghai Ranking) [10.64], the Times Higher Education (THE) ranking [10.65], the Leiden Ranking [10.66, 67], the QS ranking (2010) [10.68], the Scimago ranking [10.69], and the U-Multirank [10.70]. University rankings, though controversial, enjoy great popularity. They have become unavoidable: whether we like it or not, they are now part of academic life and of science policy. They create a reality that cannot be ignored. They are likely quite influential in guiding educational choices among young scientists and research students, and also in research collaboration strategies among universities. As soon as the first results of whatever university ranking cast their shadow, the media are happy to pounce on it. Universities use the outcomes of rankings in their rivalry with other institutions, no matter the lack of transparency of the ranking method and the significant methodological issues involved in calculating reliable ranking indicators [10.66, 71]. And when a university has fallen, say, five places, which is statistically insignificant, the university board is in trouble. It would be no exaggeration to say that the academic world, the public, and the media nowadays are obsessed by rankings.

We briefly summarize a few important issues with university rankings. Some rankings, for instance THE and QS, combine their scores for teaching and research performance into one final score. We think that this is not desirable, because teaching and research are different tasks of universities. A combined score may not take into account the university's specific mission (focus on teaching, or on research). The ARWU/Shanghai and Leiden rankings focus specifically on research only. In the ARWU/Shanghai ranking, size-dependent (e. g., number of papers in *Nature*) and size-independent measures (particularly the use of size-dependent indicators per staff member) are combined. This deserves no methodological beauty prize. If indicators are not clearly field-normalized, universities with a focus on engineering or on social sciences and humanities will be disadvantaged. The comparison of ranking scores in a time series can be seriously affected if, meanwhile, the number of universities covered by the ranking is increased substantially. Any indicator, including a ranking score, is subject to statical uncertainty. Only the Leiden Ranking uses a well-defined uncertainty measure, the stability interval.

Most rankings use citation indicators based on averages. As we discussed in Sect. 10.2, average-based indicators are sensitive to outliers. There is generally a strong correlation between average-based indicators such as the MNCS and distribution-based indicators such as the pptop10%. But there are remarkable exceptions. A famous case is the earlier-mentioned *Göttingen effect*. In 2008, a paper published by a researcher of the University of Göttingen became extremely highly cited, many thousands of times a year, within a very short time (G.M. Sheldrick, A short history of SHELX, Acta Crystallographica A 64, 112–122, 2008; cited 67 693 times (WoS Core Collection), December 1, 2018). As a result, for several years after this publication, Göttingen achieved a very high position in the rankings. But with a distribution-based indicator, the ranking position of this university was much lower. For instance, in the Leiden Ranking 2011, Göttingen ranked second based on the MNCS indicator, while it was ranked 238th based on the pptop10% indicator. Without this single extremely highly cited paper, the average-based citation impact of Göttingen would have been only half of the value, and the rank would have been 219 instead of 2—so a difference of about 200 ranking positions because of just one publication. Of course, this is an exceptional case. But there are similar less dramatic cases, and the point we make is that with exactly the same data, the difference in normalization may affect the ranking of universities with tens of positions. For a detailed discussion of the Göttingen case, we refer to [10.66].

As mentioned earlier, *retraction of papers* is an increasingly important issue, particularly in the construction of rankings. Papers are retracted for scientific misconduct such as fraud, data fabrication, plagiarism [10.162], serious errors, and sloppiness. Intentional misconduct is the main reason for retraction in about 60% of cases [10.163]. This means that bibliometric databases used for research evaluation must be continuously cleaned in order to remove retracted papers as well as all citations received by the retracted papers. This is the case for the CWTS bibliometric database, used for the Leiden Ranking. Updates on new retractions and discussions on issues related to retractions are provided, for instance, by Retraction Watch (non-undisputed, see [10.161]) and by The Scientist [10.233]. The number of retracted papers is very small: of all papers published in 2009–2013 and covered by the WoS, about 0.05% were retracted. Retractions are observed more commonly in fields with strong international competition and rapid publication processes [10.163].

For a recent comparison of five ranking systems—ARWU, Leiden, THE, QS, and U-Multirank—we refer to [10.234]. New approaches to rankings and other methods for comparison of universities from a national and international perspective, along with factors that affect the reliability of rankings, are discussed in [10.235–237].

Objections to Bibliometric Analysis

Without doubt, the application of even the most advanced bibliometric method has its limitations, and this handbook chapter discusses many of these limitations in detail. But the headaches that bibliometric analysis has caused are far from over. Next to methodological and technical problems there is also a considerable undercurrent of opinions, fears, and prejudices against bibliometric analysis. Critics maintain that the mere use of bibliometric indicators in research evaluation will lead to perverse publication behavior among authors, particularly *gaming*. Gaming is not a clear concept. It may relate to the manipulation of the data on which bibliometric indicators are based (particularly number of publications and citations) to deliberately give an inaccurate impression of research quality (negative or harmful gaming). But it also relates to seeking strategies to improve the visibility of published work, with the possible consequence of increasing the measured impact (positive or beneficial gaming). Examples of the first kind are citing your own work more frequently, arranging citation cartels, and *salami slicing*. With respect to self-citations, in advanced bibliometric methods, author self-citations are excluded in the citation counts. As a result, increased author self-citation has no effect on the value of citation impact. Arranging a citation cartel? Active researchers, and certainly research groups, receive citations from hundreds of different authors and institutions. The distribution of citations among citing authors is very skewed, and the tail of the distribution is relatively large and contributes considerably to the total number of received citations. As one might have guessed, one would need citation cartel arrangements with many institutions in order to a create a substantial increase in citation impact. Practicing salami slicing, which is dividing your work into *the smallest publishable units*? Most likely this strategy will result in a lower value of the normalized citation impact. Nevertheless, harmful gaming exists, particularly journal impact factor manipulation, for instance, by excessive journal self-citation; for a recent overview we refer to [10.206] and to the website of the Committee for Publication Ethics (COPE) [10.238]. Bibliometric researchers recently demonstrated the vulnerability of Google Scholar's citation metric to gaming by placing fake papers with many citations to their own work on a webpage of their university [10.239].

An example of the second kind of gaming is trying to publish your work in higher-impact journals. Although journal impact correlates quite significantly with the actual citation impact at the level of a research group, a study of 150 chemistry research groups [10.147] showed that top-performing groups (both based on peer judgments and as measured with the crown indicator) were on average more successful in citation impact over the entire range of journal impact. In other words, they performed better in both the lower-impact and higher-impact journals. Quality is the decisive factor, not journal impact.

Perhaps not surprisingly, other objections to bibliometric analysis, based on little more than anecdotal evidence, relate to certain persisting opinions. For example, some say that the use of bibliometric indicators in research evaluation would favor researchers who work on fashionable topics and who avoid innovative and *risky research*. I do not know of any decent empirical evidence for this assertion. And there are also *negative citations*, i. e., a publication can be criticized, for instance, because of errors and consequently be well cited. Most probably, however, erroneous publications will not be cited but rather ignored, and in the case of serious errors, the paper will be retracted. Another belief is that bibliometric indicators would encourage the writing of reviews instead of original papers, because reviews are generally more frequently cited. In advanced bibliometrics, however, the article type is taken into account in the normalization of the citation impact calculations. Yet another argument is that bibliometric analysis would encourage cannibalization of old papers by refurbishing and republishing them. Perhaps this will happen every now and then, but it is very unlikely that a research group would have a high impact over a longer period of time just by cannibalizing its older work. Nevertheless, the use of bibliometric indicators in evaluation procedures may influence the publication strategy of researchers [10.240, 241], but certainly not always in a negative sense [10.242].

A more politically and socially sensitive objection against bibliometric indicators is that they would discriminate against young researchers, women, and minorities. Indicators such as the h-index undoubtedly favor, practically by definition, the older and experienced researchers, but advanced indicators such as the normalized citation impact do not have this disadvantage. However, the supposed discrimination problem is not in the construction of the indicators as such; it lies deeper—for instance, the work of women would receive fewer citations, in cases of equal quality, compared to the work of men [10.243].

A great deal of the skepticism about bibliometric methods, for instance in the Research Excellence Framework (REF) 2014 [10.244], relates to the application of bibliometric analysis to individual papers. This is because the REF 2014 was an article-based assessment, and not an assessment of research groups or departments. Also, the Italian VQR evaluation (Valutazione della Qualità della Ricerca—Evaluation of Research Quality) [10.245] was an article-based assessment. Approximately 1000 (REF) and 10 000 (VQR) experts had to judge the quality of around 100 000–200 000 papers. The mere size of such an amount of review work per expert may raise questions about the reliability of these qualitative assessments. We think that the application of advanced bibliometric methods is optimal when applied to groups, department, or institutes, and not at the level of individual papers. The main reason is that individual papers are mostly part of a coherent research oeuvre, and thus the assessment should be applied to the oeuvre as a whole and not to isolated papers. Also, if evaluation is applied to groups, peers will most probably take important contextual information into account, such as the viability of a group and the relevance of its research [10.122]. This information is completely absent in an article-based assessment. Furthermore, if longer timescales are used, the evaluation of the entire groups' oeuvres also significantly reduces the effects of any possible gaming.

Bibliometrics is not all sorrow and misery, but in some cases critics preach hellfire and damnation in quite hysteric words, predicting that bibliometric analysis may become the ultimate tombstone of science. Even worse, bibliometric researchers should be sent to "the darkest omnivoric black hole that is known in the entire universe, in order to liberate academia forever from this pestilence" [10.246, 247]. That at least one important publication is written on a tombstone may give some comfort to the cursed bibliometricians (Fig. 10.4).

Fig. 10.4 Tombstone in the nearly 900-year-old St. Peter's Church, Leiden, of Ludolph van Ceulen (1540–1610), professor of mathematics at Leiden University, on which his calculation of the number π to 35 decimals is published (from [10.248]; Photo: Claudia Claas)

10.3.2 Real-Life Example of Evaluation Studies

Hoping that we will get over our distress, we pursue in this section the most interesting side of bibliometric indicators: their application in a real-life research evaluation. As an example, we take the bibliometric portion of the extensive evaluation of the Institute of Veterinary Research of Utrecht University (IVR-UU), which also included peer review by an international committee. Reports on both the peer review findings and the bibliometric analysis can be found in [10.249]. As we mentioned earlier, CWTS already has for decades used fully standardized procedures discussed beforehand with client institutes in order to carry out bibliometric analysis as conscientiously as possible. In this sense, advanced bibliometric analysis is practically identical to what has recently been called *contextualized scientometric analysis*.

We show in Table 10.1 the main bibliometric indicators for the entire evaluation period (2001–2011) as well as a trend analysis. There is ample empirical evidence that in the natural and life sciences, basic as well as applied, the peak number of citations occurs on average in the third or fourth year after publication. Therefore, a 4-year period is appropriate for impact assessment. The indicators *P*, MCS, MNCS, MNJS, and pptop10% and the details of the trend analysis were discussed in Sect. 10.2.

It is possible that publications are not cited within a period of 4 years, but will be cited after a longer time, i. e., the earlier-discussed phenomenon of delayed recognition or, in extreme cases, sleeping beauties [10.98, 99, 166–177]. Contrary to popular belief, sleeping beauties do not concern mostly whimsical research; about half of them are related to applied research and may be sleeping technological inventions [10.169, 171]. Delayed recognition may dilute the impact measured in a 4-year period (because they are published but not cited in the given period). This effect, however, is generally not large at the level of research groups [10.250–252]. Nevertheless, it is important to assess research performance for a longer citation window as well, and that is why the indicators are also calculated for the entire evaluation period, as can be seen in Table 10.1 and subsequent tables. Note that all indicators are corrected for self-citations. With coverage, we indicate the fraction of publications presented by the institute in its evaluation protocol that are covered by the WoS and thus also in this bibliometric analysis. We observe that the coverage of IVR-UU is about 85%, so that one can be confident not to miss a substantial part of the institute's oeuvre in the analysis.

We see that IVR-UU performs very well above the international level (MNCS = 1.58 and pptop10% = 0.18 in the last period). With an MNCS value above 1.5, such as in our example, the institute can be considered scientifically strong. The proportion of highly cited (pptop10%) papers increases from 14% to 18%. The indicator MNJS shows that the impact of the journals used by IVR-UU for their publications is significantly higher than the average of the fields to which the journals belong. In other words, IVR-UU publishes in the higher-impact journals. We also observe the lower number of publications in the last period. This is because, at the time of the analysis, papers published in 2011 did not have a full year to be cited and were excluded in this analysis. This shows that a clear explanation of the bibliometric results is necessary; otherwise the user would think that there was a significant decrease in the number of publications.

Researchers of an institute publish in journals of different fields, often in many more fields than one would tend to think. Therefore, a next step in the analysis is to construct a research profile: the *breakdown* of the institute's output into research fields. For this breakdown we use the fields as defined in the WoS (*subject areas*), but other field definitions, such as publication-level-based, are possible. The advantage of using the WoS subject areas is that many researchers are familiar with these field definitions and that they make comparison of the profiles of different groups easier. We show the research profile of IVR-UU in Fig. 10.5. Fields are ranked by their number of publications; the first 20 are shown.

The bars represent the fields, and the length of the bar is determined by the number of publications in a field. Clearly, for IVR-UU, the field of veterinary science is by far the largest. But it is also clear that IVR-UU researchers are active in many other fields, including immunology and microbiology. Furthermore, the *impact* of IVR-UU research in each of the fields is given by the MNCS values, and we see a high MNCS value for most of the fields, as is to be expected given the high MNCS of the total output. There are fields in which IVR-UU researchers have a particularly high impact, such as microbiology and food science & technology. We see also fields with a lower MNCS, for instance, immunology and endocrinology & metabolism. This is important information for the institute and its researchers in order to understand why the

Table 10.1 Bibliometric indicators for IVR-UU for the total evaluation period 2001–2011 and trend analysis in 4-year periods. This table is taken from [10.249]

IVR-UU	Coverage	P	MCS	MNCS	MNJS	pptop10%
2001–2011	*0.86*	*3891*	*13.69*	*1.46*	*1.33*	*0.17*
2001–2004	0.84	1492	6.68	1.31	1.26	0.14
2002–2005	0.85	1567	6.96	1.34	1.28	0.15
2003–2006	0.86	1542	7.30	1.38	1.30	0.16
2004–2007	0.87	1556	7.57	1.39	1.31	0.17
2005–2008	0.87	1530	7.94	1.47	1.34	0.17
2006–2009	0.88	1544	7.32	1.46	1.38	0.17
2007–2010	0.88	1626	5.84	1.56	1.42	0.18
2008–2011	0.88	1250	5.11	1.58	1.44	0.18

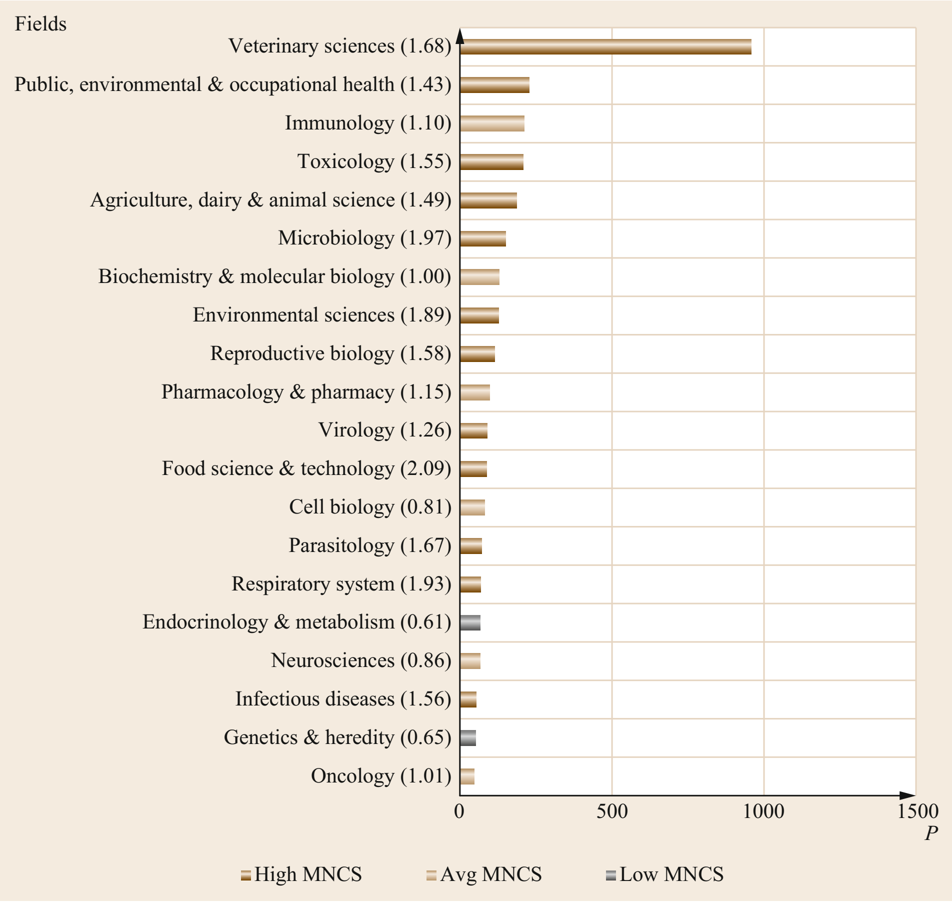


Fig. 10.5 Research profile of IVR-UU: output and impact per field, 2001–2011. A *dark brown bar* indicates a high MNCS (> 1.20), *light brown* indicates an average MNCS (between 0.80 and 1.20), and *gray bars* indicate a low (i. e., below the international average of the field) MNCS (< 0.80). MNCS values are given after the name of the field. This figure is taken from [10.249]

impact in these fields is relatively low. Similar profiles can be made for the papers *citing the work of an institute*, see for instance [10.253].

In Sect. 10.2, we have discussed indicators of collaboration. In most evaluation studies, a somewhat simpler approach is sufficient: publication with no collaboration, publications with national collaboration, and publications with international collaboration. We present the collaboration profile of IVR-UU in Fig. 10.6. In all types of collaboration, the MNCS value is significantly above world average. The findings show a quite general picture: publications in international collaboration are the largest group and have the highest impact; publications with no collaboration are the smallest group and their impact is lower (but still significantly above the worldwide average).

At the time of the evaluation, the research within IVR-UU was organized in six programs: Biology of Reproductive Cells (BRC), Tissue Repair (TR), Emotion and Cognition (EC), Risk Assessment of Toxic and Immuno-modulatory Agents (RATIA), Strategic Infection Biology (SIB), and Advances in Veterinary Medicine (AVM). As an example of bibliometric analysis at a lower aggregation level, we present results for the BRC and the TR programs. For additional results, we refer to the complete IVR-UU report [10.249]. First, we present in Table 10.2 the bibliometric indicators for both programs for the total evaluation period 2001-2011 and trend analysis in 4-year periods, similar to Table 10.1 (but without indication of coverage). The TR program started in 2006, so for this program we have data from 2006 onward.

For both programs we see high MNCS and pp-top10% values over the entire evaluation period, which confirms the high impact of the Utrecht veterinary research. In the TR program in particular, journals of high impact are used for publication (high MNJS values). The peer review judgment of the quality was "very good" for both programs [10.249]. In Fig. 10.7 we show the collaboration profiles of the BRC and TR programs. We see that in the BRC program, the publications in international collaboration receive the highest impact ratings. In the TR program, the publications in national collaboration have the highest impact, whereas

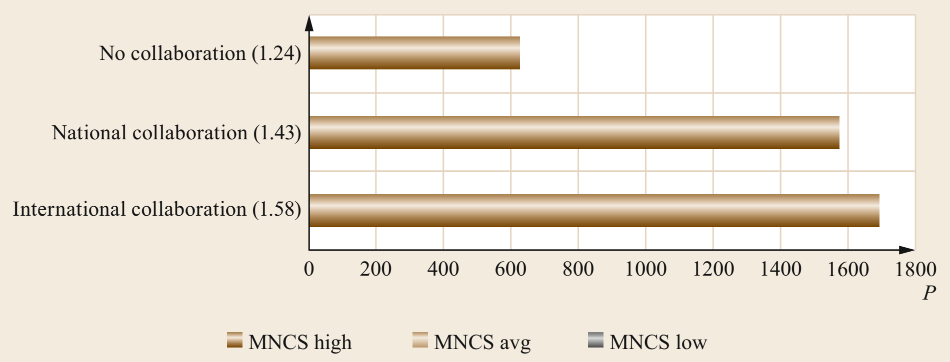


Fig. 10.6 Collaboration profile of IVR-UU: output and impact per type of collaboration, 2001–2011. A *dark brown bar* indicates a high MNCS (> 1.20), *light brown* indicates an average MNCS (between 0.80 and 1.20), and *gray bars* indicate a low (i. e., below the international average of the field) MNCS (< 0.80). MNCS values are given for each type of collaboration. This figure is taken from [10.249]

Table 10.2 Bibliometric indicators for IVR-UU programs Biology of Reproductive Cells (BRC) (*upper part*) and Tissue Repair (TR) (*lower part*) for the total evaluation period 2001–2011 and trend analysis in 4-year periods. This table is taken from [10.249]

IVR-UU/BRC	P	MCS	MNCS	MNJS	pptop10%
2001–2011	*274*	*14.39*	*1.38*	*1.14*	*0.16*
2001–2004	119	6.71	1.22	1.17	0.14
2002–2005	126	7.20	1.45	1.21	0.18
2003–2006	124	7.35	1.51	1.22	0.19
2004–2007	120	7.27	1.45	1.17	0.17
2005–2008	110	7.88	1.58	1.15	0.19
2006–2009	101	6.84	1.31	1.15	0.14
2007–2010	101	5.74	1.28	1.16	0.15
2008–2011	73	5.33	1.35	1.17	0.16
IVR-UU/TR	**P**	**MCS**	**MNCS**	**MNJS**	**pptop10%**
2006-2011	*368*	*4.97*	*1.50*	*1.46*	*0.16*
2006–2009	255	4.99	1.40	1.41	0.17
2007–2010	241	4.60	1.60	1.49	0.19
2008–2011	236	3.93	1.74	1.53	0.19

the publications with no collaboration have an average impact.

This example of *bibliometrics in practice* shows the importance of an indicator system in which all relevant indicators are calculated and presented in a coherent way, i. e., based on the same underlying data, such as a specific set of publications and a specific time period, with clear visualization of the results. A next step is automation of such an indicator system, offering the user possibilities to choose, for instance, other time periods and other target groups. This means a menu-driven application instrument for research performance assessment and monitoring of groups, university departments, institutes, and research programs. Novel features can be added, such as geographical maps with indication of research groups worldwide citing and/or collaborating with the institutes or within research programs. In this way, crucial information on the impact and other properties of research can be provided. Such an automated indicator system for monitoring and evaluation based on the methodological principles discussed in this chapter has recently become available in a first version, with new extensions currently in preparation [10.254]. It hardly comes as a surprise that this type of advanced bibliometric analysis offers the user information on research performance of a scope and quality far beyond the level of what we described as easy-to-go bibliometrics.

What can we conclude from all this? Above all, that it is important to have diverse indicators, as discussed in the above example. It may help to understand peculiarities of the assessed groups. This diversity can be extended with several other bibliometric indicators, for instance the percentage of not-cited papers and the number of authors (more authors may affect impact [10.255]), as well as with *non-bibliometric indicators* such as properties of the most highly cited papers of the group (for instance: are these reviews, or original research work?), patents [10.256, 257], acknowledgments, altmetrics data, staff size and staff composition of the research group, and other input measures [10.137]. For an example of a university-wide evaluation in which both bibliometric analysis and peer review are used, we refer to the evaluation study for Uppsala University [10.123].

10.3.3 Summary of Guidelines for the Use of Bibliometric Indicators

Number of publications is database-dependent and language-dependent, and there are large field-specific

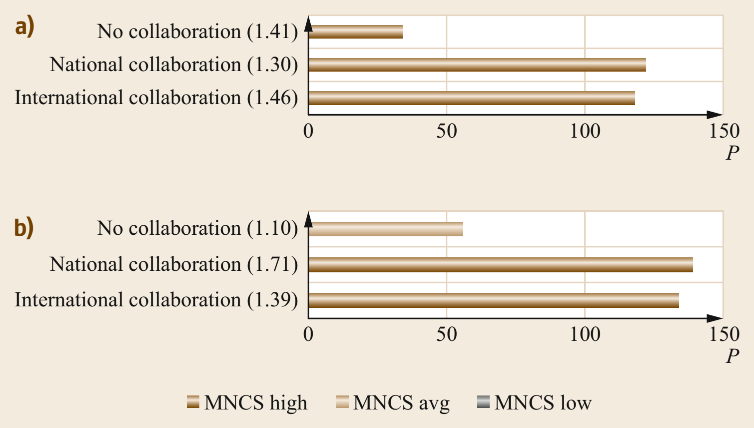


Fig. 10.7 **(a)** Collaboration profile of the IVR-UU program Biology of Reproductive Cells. **(b)** Collaboration profile of the IVR-UU program Tissue Repair. MNCS values are given for each type of collaboration. This figure is taken from [10.249]

differences in publication culture. This is particularly the case for most of the humanities and some of the social sciences as well as the engineering fields. Nevertheless, in the humanities, the social sciences and the engineering fields, journal publications become more important and bibliometric indicators can be applied experimentally. Be aware that not all publications of a research group will be captured—for example, publications written for practitioners. Thus, the productivity of a research group cannot be based on the number of papers covered by the WoS or Scopus alone. Specify the database used, and do not compare different fields in terms of absolute number of publications. Fractional counting has the advantage that averages add up worldwide to 1, but it is often not fair in fields where collaboration is practically inevitable, such as astronomy. Thus, both fractional and whole counting are necessary. The difference between the two counting modalities provides information on the robustness of the whole indicator system: if the difference between the modalities is large, then there is good reason to analyze the data in greater detail to see what is going on. In international comparisons, it is necessary to remove the non-English-language journal papers. This may differ by 20–25% in the impact for French and German research groups and institutes. It is also necessary to remove retracted papers and all citations received by these papers. Attribution of publications to affiliations is a difficult and often underestimated problem. The bibliometric researchers must have the expertise and capacity to clean, unify, and otherwise define possible research affiliations as best as possible. Moreover, research is often teamwork, and the lifetime of a team may be shorter than the period studied. Take into account composition of and changes in teams. Crucially important: always use a set of consistent indicators, such as the set P, MNJS, MNCS, MCS, and pptopx% discussed in this chapter.

Number of citations, i. e., the number of citing publications, is also database-dependent, language-dependent, and directly related to the publication culture; there are large field-specific differences in citation culture. Again, this is particularly the case for most of the humanities and some of the social sciences as well as the engineering fields. The building blocks for citation analysis are pairs of citing and cited publications. We must have very accurate knowledge of the citation data on this level. We can then aggregate to higher levels such as groups, institutes, programs, and universities. Specify precisely the database used; do not compare different fields in terms of absolute numbers of citations. Carefully define citation windows, for instance with the roof-tile procedure. The choice of citation window length and structure of the citation window is critical: the importance of a publication does not necessarily appear immediately, even to peers, and identification of quality may take considerable time. Therefore, it is important to make assessments with shorter as well as longer time windows. The number of citations can be affected by type of collaboration and number of authors.

Number of highly cited publications requires an accurate calculation of the entire citation distribution within a field; this is necessary to determine, for instance, the top-1%, top-5%, and top-10% publications; the distribution function depends on the definition of the field. Again, choice of citation window length and structure of the citation window are essential. Specify precisely how the fields are defined, in which database, and how the citation distribution is calculated. The top-10% category is generally more informative than the top 1% or top 5%. A top-1% publication is not necessarily a breakthrough or highly innovative paper. Papers in which extended reviews or important established research methods and algorithms are presented can be very highly cited. Such papers are important, but they are often mainstream and not necessarily innovative. Also, highly cited papers could be work that fits well in a specific trend, which may be short-lasting.

Number of non-cited publications depends on the choice of citation window length and structure of the citation window; as we already noted, in the case of delayed recognition and sleeping beauties, not being cited is a time-dependent phenomenon. Thus, specify in which period of time the number of non-cited publications is calculated.

Research field classification by journals is problematic because many important journals are too broad to be classified in one field. Therefore, journal-based classification is less accurate than publication-based classification. Nevertheless, a journal classification such as the WoS journal categories is still useful because it is a simple, easy-to-understand, well-established classification known in the academic community worldwide, and it is quite stable. We stress that in evaluation studies, it is preferable to use a publication-based classification next to the journal-based classification. In an advanced bibliometric data system with proper algorithms, this can be done without too much trouble.

Normalized citations are very sensitive to the definition of a field. For instance, the MNCS or pptop10% indicators will have different values if the normalization is based on a publication-level-defined field as compared to normalization based on a WoS field. Choice of citation window length and structure of the citation window is critical. Specify precisely how the fields are defined, in which database, and how the normalization is calculated, for instance on the basis of averages or on the basis of the citation distribution. Calculate stability intervals or other proper measures of statistical significance. Show that the applied normalization is a proper one, given the often large heterogeneity of citation density particularly within fields. Also, journal impact has to be normalized in a similar way as the normalization of citations to the publications of a research group. Combine information on output and impact by field in a research profile as discussed in this chapter.

National and international collaboration measures are database-dependent, and there are large field-specific and country-specific differences in collaboration culture. Also, large differences exist between fields for the fraction of co-authored publications and number of co-authors per publication. We already stressed that definition of (collaborating) affiliations is a cumbersome task. In order to visualize collaboration patterns, apply advanced network/mapping techniques combined with citation-impact analysis as described in this chapter, for instance with help of the VOSviewer. Specify precisely the database used. Differences between fields provide important information on collaboration cultures. Impact calculations for different types (e. g., national, international) of collaboration are necessary. Use carefully designed and transparently calculated collaboration indicators such as those described in this chapter. Impact calculations for different numbers and percentages of co-authorship may provide important information about teamwork.

Number of patents is characterized by large differences between scientific fields in patenting possibilities as well as large differences between technological sectors in patenting culture. There are different classification systems with different fine structures. Classification of technological products within sectors is necessary, and the rules as to how smaller products are considered to be components or parts of a larger product may differ considerably between technological sectors. Specify precisely how the technological sectors are defined, in which database. Patent-to-patent and paper-to-patent citations may reveal important information on the interaction between science—and especially applied research—and technological developments.

Altmetrics indicators tend to measure the social impact of research rather than the scientific impact. There are large differences in the extent to which specific developments in science are trendy for the public. Trendiness changes continuously over time depending on current social, economic, political, and cultural issues. In addition, critical thinking is the fundament of science, often in disagreement with majority thinking as reflected by trends. Field-specific normalization of altmetric indicators is extremely difficult, if not impossible. Regard altmetrics as an experimental approach to capture social relevance.

10.4 What Is a Bibliometric Science Map?

After having reveled long enough in the Wild West of bibliometric indicators, it is now time to explore the other side of the bibliometric world, science mapping. It is a challenge to identify hidden patterns in the enormous amount of published scientific knowledge, given that all these publications (and patents) are connected by common references and concepts. Co-citation and co-word techniques are important examples of approaches for unraveling this gigantic network of interrelated pieces of scientific knowledge. These are important steps toward *imaging cognitive processes*. Maps of science, with the locations of the major actors, are specific representations of scientific activities. They have practical values (e. g., strategic overviews) as well as cognitive values (e. g., what type of scientific activities are represented on the map, and how they develop over time).

10.4.1 Basics and the Construction of Science Maps

We begin by recalling that the bibliometric network PAN forms the basis for the construction of bibliometric indicators as well as for science mapping. If the publication attributes are cited papers, the bibliometric tool *CitNetExplorer* can be applied to map the citation relations between the target paper and its cited papers (references) [10.110], including a timescale. This citation-links mapping enables us to find the *scientific roots* of the target paper, and possibly an older but important breakthrough paper. With the CitNetExplorer, the target paper can also be mapped with its citing publications, as we will see further on. For a more detailed analysis and mapping of the publications citing a target paper, a second bibliometric tool, the *VOSviewer*, can be applied [10.113]. As explained in Sect. 10.2, BC networks and CC networks can be constructed. In the BC network, *the citing publications* are mapped on the basis of the co-occurrences of their references. With advanced clustering techniques, a BC network transforms into a map which visualizes a structured landscape of all publications citing a specific target paper. It thus provides us with information on the research building on the target paper.

In the CC network, *the references of the citing papers* are mapped. Again, with advanced clustering techniques (related to modularity-based clustering), a CC network transforms into a map which visualizes a structured landscape of the references of the citing publications. The target paper is the reference of all citing papers. So it will take a central position in the CC map. This provides us with information as to which publications are often cited together with the target paper, and therefore publications that are probably just as important as, or even more important than, the target paper. As an example, we take the highly cited paper of this author on rankings [10.71]. We download the WoS record of this paper with its cited papers (the references) as well as its citing papers. In a next step, these data are uploaded into the CitNetExplorer. Fig. 10.8 presents the results of the CitNetExplorer application. The upper part of the figure shows the papers *cited by the target paper*; the lower part represents the papers *citing the target paper*. In both cases we marked the target paper with a square in the figure. Connecting lines indicate citation relations. These lines always go in an upward direction, which is backward in time.

These citing papers do, of course, cite additional papers, including mutual citations within the set of citing papers. These mutual connections (number depends on the citation threshold) are also visible in Fig. 10.8a. With the uploaded set of citing papers, a more comprehensive analysis of all citation links can be performed interactively with the CitNetExplorer by *tuning the citation threshold*. This automation of reference analysis has promising applications, given possibly hidden patterns within references [10.258, 259].

As a next step, we upload the same data of the target paper in the VOSViewer. In Fig. 10.9 we show the CC map of the *papers citing the target paper*. This creates a visualization, including clustering, of the research used (cited) by these citing papers. The target paper is in a central position because, by definition, all of the citing papers have the target papers as one of their references. Co-citation may involve more than just two papers, and thus one can expect clustering of papers that depends on the reference preferences of the citing papers. Given the heterogeneity of the citing papers, different clusters will exist, as can be seen in the figure, where clusters are distinguished by color. Also, the VOSViewer makes it possible to create maps with different thresholds: *tunable co-citation analysis*. At a high threshold, only the strongest co-citation relations will be mapped. Thus, tunable analysis increases the value of interactive mapping, but at the same time we caution for an overload of information.

The above-discussed results are all based on citation links—in other words, on the PAN in which the attributes are cited papers. If the attributes are *concepts*, mapping procedures similar to those for citations can be carried out. To do so, we can apply natural language processing (text mining) to extract the important, publication-specific concepts (terms such as keywords or noun phrases) from the titles and abstracts—or even from the full text—of a set of publications. Alternatively, keywords given by the authors and by the database can be used. Choice and selection of relevant keywords is not a trivial thing, and it is important to experiment with and compare different methods [10.260]. By measuring all co-occurrences of any possible pair of concepts, co-word maps (also called term maps) can be created, in which the conceptual structure of the research represented by the set of publications is visualized. Thus, on the basis of the above considerations, a term map is a two-dimensional representation of a field in which strongly related terms are located close to each other, and less strongly related terms are located further away from each other.

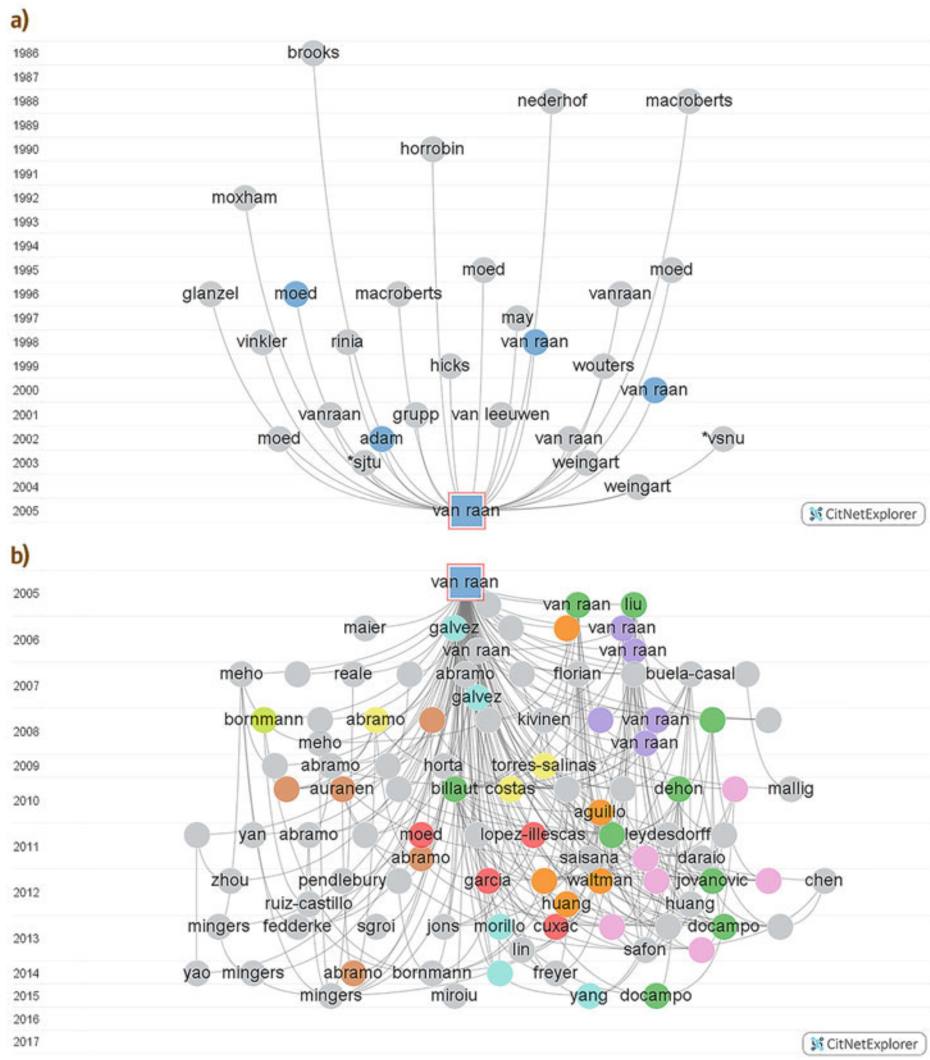


Fig. 10.8a,b Maps of the citation links of the target paper. **(a)** The cited papers (references) of the target paper; **(b)** the citing papers of the target paper. *Connecting lines* indicate citation relations; these go always in an upward direction. *Colors* indicate clusters based on mutual citation relations

A term map provides an overview of the structure of a field. Different areas in a map correspond to different subfields or research areas. A time series of maps enables us to discover emerging and converging research themes.

In Fig. 10.10 we present the term map based on author—as well as database—given keywords of all papers citing our target paper. We clearly observe many concepts directly related to the target paper, particularly *ranking* and *universities*. Colors indicate clusters of concepts, and these clusters can be seen as research themes. For instance, the red cluster is about methods and properties of bibliometric indicators, the purple cluster is about applications of bibliometric indicators, the green cluster is about the position of universities in rankings, and the light green cluster focuses on problems in the ranking methodology. It is important to note that in the interpretation of a term map, only the distances between terms are relevant. A map can be freely rotated, because this does not affect inter-term distances. This also implies that the horizontal and vertical axes have no special meaning.

There is a continuous effort to further improve mapping of science and its visualization. In recent work, multilayer concept networks are studied [10.261], and comparisons are made between bibliometric maps and topic modeling [10.262, 263]. A combination of different types of maps based on journal-to-journal citations, shared author keywords, and co-occurrence of title words and cited reference is discussed in [10.264]. A quite rarely conducted but interesting experiment is *back-to-the-future mapping*, i. e., reconstructing the

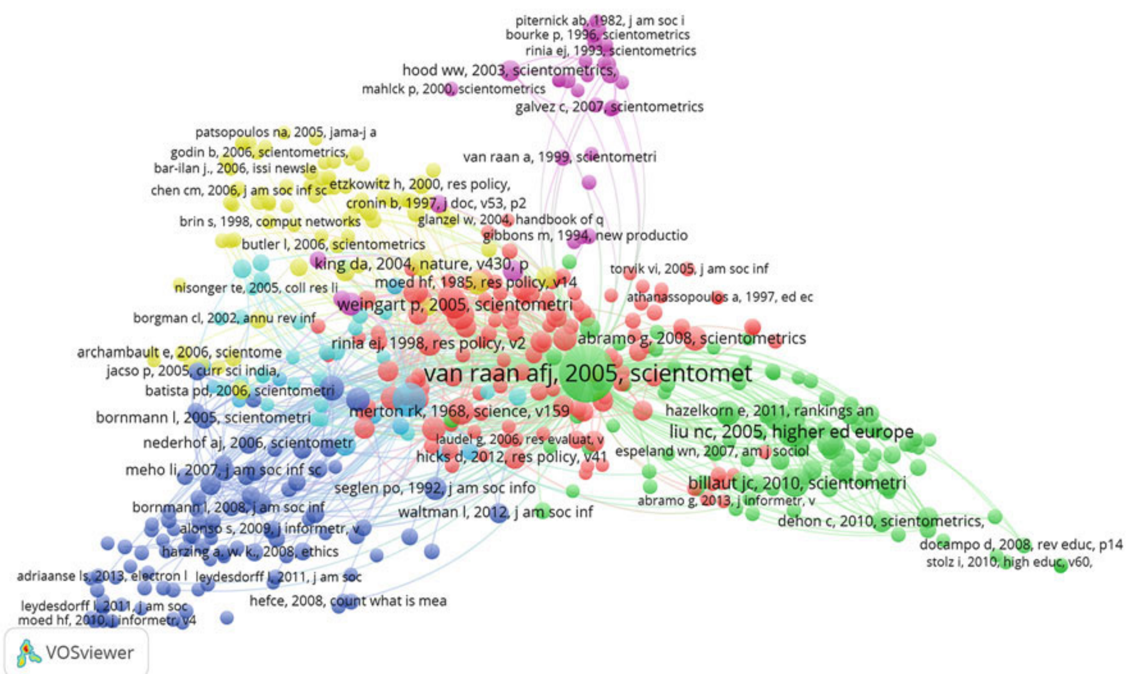


Fig. 10.9 Co-citation map of papers citing the target paper (co-citation threshold = 3). The size of the *circles* is proportional to the number of times a paper is cited in the uploaded set. By definition, the target paper [10.71] is the most cited paper, as all papers in the set cite the target paper

past (maps of research fields, say, 10 years ago) with currently used concepts, and the other way around, the construction of current research with concepts used 10 years ago [10.265].

10.4.2 Combining Citation Analysis and Science Mapping

A next step is the combination of the two main bibliometric methods, citation analysis and mapping. We present in Fig. 10.11 as an example (from [10.266]) the medical field of clinical neurology defined by its journal category in the WoS (for further work on science mapping based on WoS categories we refer to [10.267, 268]). The map is based on all publications classified as *article* or *review* and published between 2006 and 2010 (105 405 in total). For each publication, citations are counted through the end of 2011. Each circle represents a term. In this case, terms have been extracted by text mining from the titles and abstracts of the publications in a main field. For some terms, only a circle is displayed, not the term itself. This is done in order to prevent terms from overlapping each other. The size of a term (circle) reflects the number of publications in which the term occurs. Larger terms occur in a greater number of publications. As discussed above, the distance between two terms reflects their relatedness. In general, the closer the distance between two terms, the stronger their relatedness, as measured by the number of publications in which these terms occur together.

The relative citation impact of each term is determined and indicated by a color. In order to correct for the age of a publication, for each publication the number of citations is divided by the average number of citations of all publications that appeared in the same year. This yields a time-consistent normalized citation score for a publication. A score of 1 means that the number of citations of a publication equals the average of all publications that appeared in the same field and in the same year. Next, for each of the 2000 terms, the normalized citation scores of all publications in which the term occurs (in the title or abstract) are averaged. The color of a term is determined by the resulting average score. Colors range from dark blue (average score of 0) to green (average score of 1) to red (average score of 2 or higher). Hence, a blue term indicates that the publications in which a term occurs have a low average citation impact, while a red term indicates that the underlying publications have a high average citation impact. VOSviewer software is used to visualize the term map resulting from the above steps.

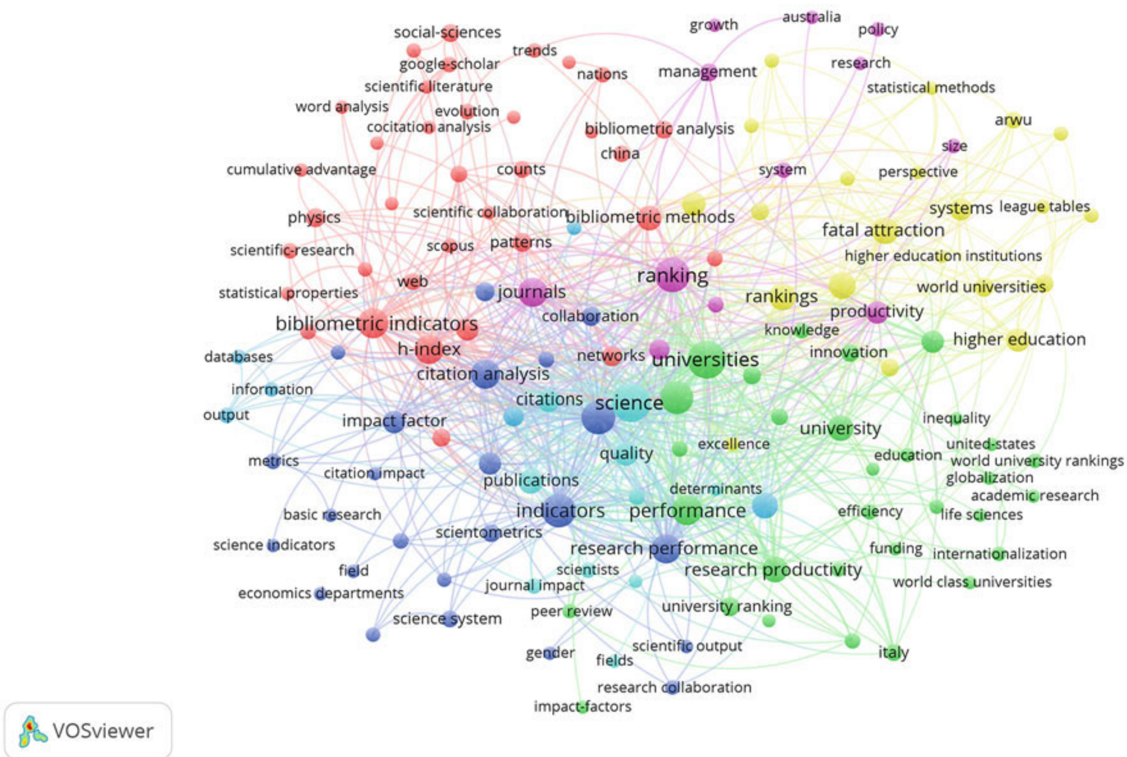


Fig. 10.10 Concept (term) map of the papers citing the target paper (mapping parameter: co-occurrence threshold = 3)

Look at the division of impact over the map: we observe a striking feature which is common to most medical fields. The typical *clinical* (hospital-related) research areas tend to be located mainly in the left part of the maps, and *basic* research areas mainly in the right part. Connections between basic research areas and clinical research areas are also visible. The maps display bridges that potentially represent *translational research*, that is, research aimed at translating basic research results into clinical practice. The color difference between the left and the right part is striking. Thus, the distinction between different research areas is visible in the structure of the maps, but also in the impact-related colors of the terms. In general, clinical research has a below-average impact (blue color), and the research areas that are more focused on basic research have an above-average citation impact (red color). We note that within an area in a map, terms are usually colored in a quite consistent way: terms tend to be surrounded mainly by other terms with a similar color. This is an important indication of the robustness of the maps.

In this section, we discussed how science mapping as an advanced visualization technique based on bibliometric data provides researchers and research institutions with an overview of the structure of a research field. It is important to note that term maps can be constructed with *any set* of chosen publications, and not necessarily with a predefined set such as the WoS or Scopus journal-based fields. For instance, maps can be made on based on the publication-level clusters discussed in the preceding sections. But one can also create sets of publications that are defined by research programs of institutes and organizations, or the publications of a university department, and use these sets to make maps of their research. Mapping of publication-level clusters based on direct citations can be used to discover emerging fields [10.269], particularly by identifying clusters with no publications in the first few years, followed by a rapid increase in citations in the subsequent years. Author co-citation analysis is also applied to detect emerging fields [10.270]. Recent work has discussed the use of heterogeneous bibliographic networks as an efficient tool for identifying emerging topics [10.271].

Another important application is the use of mapping to visualize research related to a specific socioeconomic problem. This approach enables institutes and organizations to assess their societal impact by determining how and where their research is involved in tackling

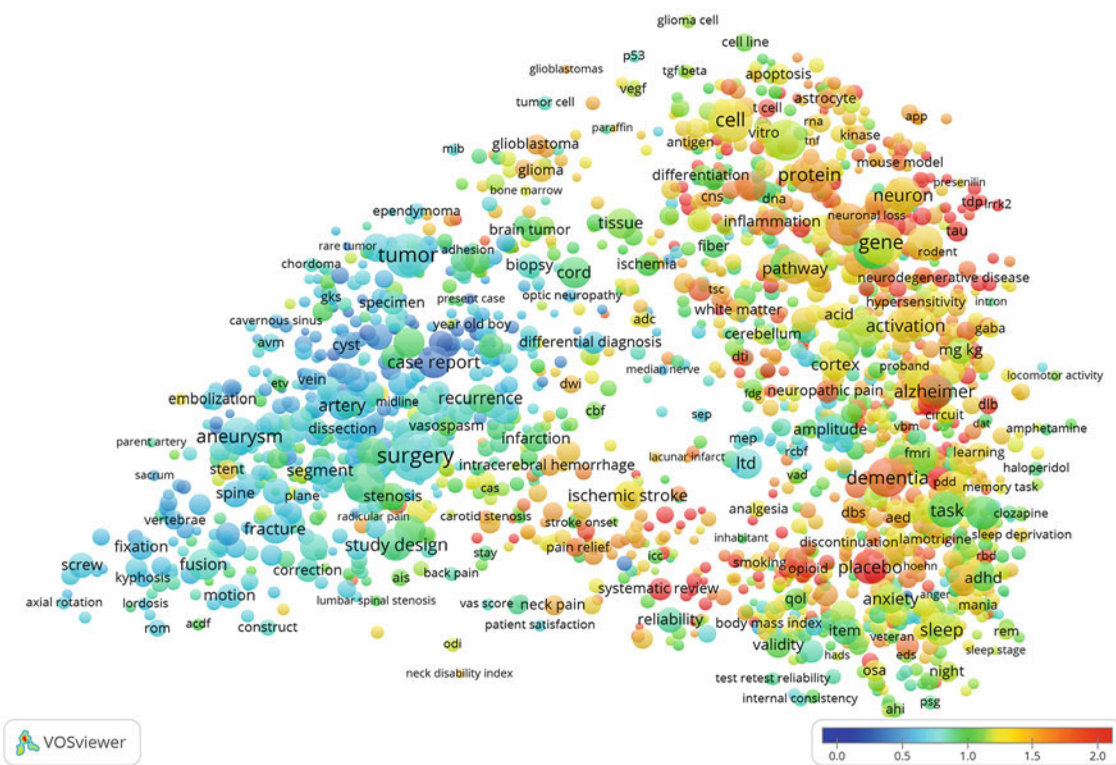


Fig. 10.11 Term map of neurology. *Colors* indicate local citation density. For a detailed discussion of this map we refer to [10.266]

a specific socioeconomic problem. A broader approach could be the following. Most likely there will be quite a strong societal consensus in defining the, say, ten most important socioeconomic problems. Identify all publications related to these problems, for instance, by using relevant concepts. Next, create maps for these problems, which are now defined as research fields based on the identified publications. Finally, use the maps to determine where specific groups, institutes, universities, or countries are active, as well as their impact in each of the socioeconomic fields.

Our discussion of science maps hopefully makes clear that mapping is a fascinating part of the advanced bibliometric methodology. It is constantly in a stage of further development, but as we showed, challenging applications are already within reach.

10.5 Can Science Be Measured?

Forty years ago, *Gerald Holton* [10.272] posed this question. The question is quite topical, and related to theories of scientific development. It is an everlasting endeavor in the social studies of science to find a final theory of science which provides *the* methodology of measurement. With very high probability, such a theory doesn't exist. It is normal practice in empirical science to begin investigations without theoretical clarification, and then later try to establish a model to explain the findings. Certainly, in such measurements, we do have at least implicit basic ideas about *how things work*, and the same is true for the construction and application of the indicators discussed in this chapter. Therefore, it is crucial to make these implicit assumptions clear to the outside world. But we need more empirical work based on the richness of available and future data in order to develop a better quantitative understanding of the processes underlying scientific evolution. Take, for example, science maps: do science maps derived from citation- or concept-similarity data have reality in a strict spatial sense? In other words, do measures of similarity imply the existence of metric space? For instance, it remains fascinating that science can be represented quite well in two-dimensional space. Nevertheless, maps based on exactly the same set of publications but constructed with different types of concepts, particularly text-mined concepts versus author-given concepts, can differ substantially. Changing thresholds for the occurrence as well as the co-occurrence of concepts will increase the diversity of maps even more, still based on exactly the same set of publications. So we need further research to determine what the best achievable *cartographic representation* of scientific research is, and why.

In this handbook chapter, we showed that the network of publications, with their attributes such as references (cited papers) and concepts, provides necessary elements to develop an advanced bibliometric methodology for measuring and mapping important aspects of science. We discussed how these measurements demonstrate the international influence of scientific work in a reliable, transparent, and objective way, particularly in the natural science and medical fields. Based on the same network principles, the advanced bibliometric methodology allows us to create science maps and to discover patterns in the structure and evolution of fields. Bibliometric mapping makes it possible to identify interdisciplinary developments and emerging fields, knowledge flows between fields of science, and research related to important socioeconomic issues.

Our choice was to focus on the above issues; other important applications of advanced methodology were mentioned only briefly, with relevant references. Many of these other applications relate directly to challenging topics that are open to further research: statistical properties of the publication network, its changes over time, improved clustering, and community detection; aging and durability of scientific literature and delayed recognition; growth and fragmentation of science; collaboration, mobility, and scientific migration flows; the interface between science and technology, particularly the role of new devices, machines, instruments; the combination of bibliometric and non-bibliometric data, particularly patent data; mapping with full-text analysis; the role of women in science; the role of cities in knowledge production and innovation.

The above shows that bibliometric indicators and maps are not just evaluation tools for science policymakers, research managers, and individual researchers, but also powerful measurement instruments in the study of science. Any instrument inevitably gives partial and distorted *images of reality*, caused by the characteristics of the instrument. Therefore, it is critically important in advanced bibliometric research, as in all empirical fields of science, to know the characteristics of instruments. This chapter is an attempt to increase this knowledge and to encourage further exploration of the quantitative properties of science.

Acknowledgments. The author thanks Nees-Jan van Eck for the construction of the publication-level citation-based network map and the citation density neurology map. The text parts on the methodology of the Leiden Ranking, the CitNetExplorer, and the VOSviewer are largely based on the descriptions in the relevant CWTS webpages by Ludo Waltman and Nees-Jan van Eck.

References

10.1 A.F.J. van Raan: Measuring science. Capita selecta of current main issues. In: *Handbook of Quantitative Science and Technology Research*, ed. by H.F. Moed, W. Glänzel, U. Schmoch (Springer, Dordrecht 2004) pp. 19–50

10.2 A.F.J. van Raan: Citations, h-index, journal impact and rankings: Not all sorrow and misery. CWTS: A short history of measuring science. In: *Perspectives on the Past: 50 Years of FSW*, ed. by J. van Holsteyn, R. Mom, I. Smit, H. Tromp, G. Wolters (Biblioscope, Utrecht 2013) pp. 86–103

10.3 A.F.J. van Raan: Advances in bibliometric analysis: research performance assessment and science mapping. In: *Bibliometrics: Use and Abuse in the Review of Research Performance*, Wenner-Gren International Series, Vol. 87, ed. by W. Blockmans, L. Engwall, D. Weaire (London, Portland 2014) pp. 17–28

10.4 F. Narin: *Evaluative Bibliometrics: The Use of Publication and Citation Analysis in the Evaluation of Scientific Activity* (National Science Foundation, Washington D.C. 1976)

10.5 A. de Candolle: *Histoire des sciences et des savants depuis deux siècles* (Fayard, Paris 1987), 1st edn. 1873, 2nd edn. 1885 (Genève/Basel, H. Georg)

10.6 A.J. Lotka: The frequency distribution of scientific productivity, J. Wash. Acad. Sci. **16**, 317–323 (1926)

10.7 S.C. Bradford: *Documentation* (Crosby, London 1948)

10.8 R.S. Daniel, C.M. Louttit: *Professional Problems in Psychology* (Prentice-Hall, New York 1953)

10.9 E. Garfield: Citation indexes for science: A new dimension in documentation through association of ideas, Science **122**(3159), 108–111 (1955)

10.10 P.F. Wouters: *The Citation Culture*, Ph.D. Thesis (University of Amsterdam, Amsterdam 1999)

10.11 D.J. de Solla Price: *Little Science, Big Science* (Yale Univ. Press, New Haven 1963)
10.12 D.J. de Solla Price: Networks of scientific papers, Science **149**, 510–515 (1965)
10.13 R.K. Merton: The Matthew effect in science. The reward and communication systems of science are considered, Science **159**(3810), 56–63 (1968)
10.14 D.J. de Solla Price: *Science Since Babylon* (Yale Univ. Press, New Haven 1975)
10.15 D.J. de Solla Price: A general theory of bibliometric and other cumulative advantage processes, J. Am. Soc. Inf. Sci. **27**(5/6), 292–306 (1976)
10.16 M.M. Kessler: Bibliographic coupling between scientific papers, Am. Doc. **14**, 10–25 (1963)
10.17 H. Small: Co-citation in the scientific literature: a new measure of the relationship between two documents, J. Am. Soc. Inf. Sci. Technol. **24**, 265–269 (1973)
10.18 S. Cole, J.R. Cole: Scientific output and recognition, Am. Sociol. Rev. **62**, 377–390 (1967)
10.19 J.R. Cole, S. Cole: The Ortega hypothesis, Science **178**, 368 (1972)
10.20 National Science Board: *Science Indicators 1972* (Government Printing Office, Washington DC 1973)
10.21 OECD: *The Measurement of Scientific and Technological Activities. Frascati Manual* (Organization for Economic Co-operation and Development (OECD), Paris 1963)
10.22 E. Garfield: Citation analysis as a tool in journal evaluation, Science **178**(4060), 471–479 (1972)
10.23 E. Garfield: *Citation Indexing: Its Theory and Application in Science, Technology and Humanities* (Wiley, New York 1979)
10.24 G. Pinski, F. Narin: Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics, Inf. Process. Manag. **12**(5), 297–312 (1976)
10.25 M.P. Carpenter, M. Cooper, F. Narin: Linkage between basic research literature and patents, Res. Manag. **23**(2), 30–35 (1980)
10.26 K.H. Chang: *Evaluation and Survey of a Subfield of Physics: Magnetic Resonance and Relaxation Studies in the Netherlands*, FOM-Report, Vol. 37175 (FOM, Utrecht 1975)
10.27 C.J.G. Bakker: *Elektronenmicroscopie in Nederland*, FOM-Report, Vol. 43105 (FOM, Utrecht 1977)
10.28 Y. Elkana, J. Lederberg, R.K. Merton, A. Thackray, H. Zuckerman (Eds.): *Toward a Metric of Science: The Advent of Science Indicators* (Wiley, New York 1978)
10.29 D.D. Beaver, R. Rosen: Studies in scientific collaboration, 1: Professional origins of scientific coauthorship, Scientometrics **1**, 65–84 (1978)
10.30 G.N. Gilbert: Measuring the growth of science-review of indicators of scientific growth, Scientometrics **1**, 9–34 (1978)
10.31 M.J. Moravcsik: Phenomenology and models of growth of science, Res. Policy **4**, 80–86 (1975)
10.32 J. Vláchy: Mobility in science. Bibliography of scientific career migration, field mobility, international academic circulation and brain drain, Scientometrics **1**, 201–228 (1979)
10.33 H.F. Moed, W.J.M. Burger, J.G. Frankfort, A.F.J. van Raan: The use of bibliometric data for the measurement of university-research performance, Res. Policy **14**(3), 131–149 (1985)
10.34 B.R. Martin, J. Irvine: Assessing basic research: Some partial indicators of scientific progress in radio astronomy, Res. Policy **12**, 61–90 (1983)
10.35 D. Sullivan, D. Koester, D.H. White, R. Kern: Understanding rapid theoretical change in particle physics- a month-by-month co-citation analysis, Scientometrics **2**, 309–319 (1980)
10.36 H.D. White, B.C. Griffith: Author cocitation – a literature measure of intellectual structure, J. Am. Soc. Inf. Sci. **32**, 163–171 (1981)
10.37 K.W. McCain: Longitudinal author cocitation mapping – the changing structure of macroeconomics, J. Am. Soc. Inf. Sci. **35**, 351–359 (1984)
10.38 S.D. Haitun: Stationary scientometric distributions. 1: Different approximations, Scientometrics **4**, 89–104 (1982)
10.39 A. Schubert, W. Glänzel: Statistical reliability of comparisons based on the citation impact of scientometric publications, Scientometrics **5**, 59–74 (1983)
10.40 B.C. Peritz: A classification of citation roles for the social sciences and related fields, Scientometrics **5**, 303–312 (1983)
10.41 M.E.D. Koenig: Bibliometric indicators versus expert opinion in assessing research performance, J. Am. Soc. Inf. Sci. **34**, 136–145 (1983)
10.42 A.L. Porter, D.E. Chubin: An indicator of cross-disciplinary research, Scientometrics **8**, 161–176 (1985)
10.43 A.G. Heffner: Funded research, multiple authorship, and sub-authorship collaboration in 4 disciplines, Scientometrics **3**(1), 5–12 (1981)
10.44 C. Fenichel: Online searching – measures that discriminate among users with different types of experiences, J. Am. Soc. Inf. Sci. **32**(1), 23–32 (1981)
10.45 K. Pavitt: Patent statistics as indicators of innovative activities—possibilities and problems, Scientometrics **7**(1/2), 77–99 (1985)
10.46 L.G. Soete, S.M.E. Wyatt: The use of foreign patenting as an internationally comparable science and technology output indicator, Scientometrics **5**(1), 31–54 (1983)
10.47 M. Callon, S. Bauin, J.P. Courtial, W. Turner: From translation to problematic networks: An introduction to co-word analysis, Soc. Sci. Inf. **22**, 191–235 (1983)
10.48 A. Rip, J.P. Courtial: Co-word maps of biotechnology—an example of cognitive scientometrics, Scientometrics **6**, 381–400 (1984)
10.49 A.F.J. van Raan (Ed.): *Handbook of Quantitative Studies of Science and Technology* (North Holland, Amsterdam 1988)
10.50 A. Schubert, T. Braun: Relative indicators and relational charts for comparative-assessment of publication output and citation impact, Scientometrics **9**(5-6), 281–291 (1986)
10.51 T. Braun, W. Glänzel, A. Schubert: World flash on basic research—the newest version of the facts

and figures on publication output and relative citation impact of 100 countries 1981–1985, Scientometrics **13**(5-6), 181–188 (1988)
10.52 R.R. Braam, H.F. Moed, A.F.J. Van Raan: Mapping of science by combined co-citation and word analysis, I: Structural aspects, J. Am. Soc. Inf. Sci. **42**, 233–251 (1991)
10.53 R.R. Braam, H.F. Moed, A.F.J. Van Raan: Mapping of science by combined co-citation and word analysis, II: Dynamical aspects, J. Am. Soc. Inf. Sci. **42**, 252–266 (1991)
10.54 P.O. Seglen: The skewness of science, J. Am. Soc. Inf. Sci. **43**, 628–638 (1992)
10.55 S. Arunachalam, R. Srinivasan, V. Raman: International collaboration in science—participation by the Asian giants, Scientometrics **30**, 7–22 (1994)
10.56 F. Narin: Patent bibliometrics, Scientometrics **30**, 147–155 (1994)
10.57 F. Narin, K.S. Hamilton, D. Olivastro: The increasing linkage between US technology and public science, Res. Policy **26**, 317–330 (1997)
10.58 U. Schmoch: Tracing the knowledge transfer from science to technology as reflected in patent indicators, Scientometrics **26**, 193–211 (1993)
10.59 A.F.J. van Raan: Fractal dimension of co-citations, Nature **347**, 626 (1990)
10.60 P.O. Seglen: Causal relationship between article citedness and journal impact, J. Am. Soc. Inf. Sci. **45**, 1–11 (1994)
10.61 H.F. Moed, T.N. van Leeuwen: Improving the accuracy of the Institute for Scientific Information's journal impact factors, J. Am. Soc. Inf. Sci. **46**, 461–467 (1995)
10.62 H.F. Moed, T.N. van Leeuwen: Impact factors can mislead, Nature **381**, 186 (1996)
10.63 A.F.J. van Raan, T.N. van Leeuwen: *Identifying the Fields for Mapping RTD Excellence in Life Sciences. First Approach* (European Commission, Brussels 2001), Contract COPO-CT-2001-00001
10.64 Academic Ranking of World Universities: http://www.shanghairanking.com
10.65 Times Higher Education World University Rankings: https://www.timeshighereducation.com/world-university-rankings
10.66 L. Waltman, C. Calero-Medina, J. Kosten, E.C.M. Noyons, R.J.W. Tijssen, N.J. van Eck, T.N. van Leeuwen, A.F.J. van Raan, M.S. Visser, P. Wouters: The Leiden Ranking 2011/2012: data collection, indicators, and interpretation, J. Am. Soc. Inf. Sci. Technol. **63**(12), 2419–2432 (2012)
10.67 Leiden Ranking 2016: http://www.leidenranking.com/information/indicators
10.68 QS Top Universities: https://www.topuniversities.com/university-rankings
10.69 Scimago Institutions Rankings: http://www.scimagoir.com
10.70 U-Multirank: http://www.umultirank.org
10.71 A.F.J. van Raan: Fatal Attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods, Scientometrics **62**(1), 133–143 (2005)
10.72 J.E. Hirsch: An index to quantify an individual's scientific research output, Proc. Natl. Acad. Sci. USA **102**(467), 16569–16572 (2005)
10.73 A.W. Harzing: *The Publish or Perish Book. Your Guide to Effective and Responsible Citation Analysis* (Tarma Software Research, Melbourne 2010)
10.74 A.W. Harzing: Publish or Perish, http://harzing.com/resources/publish-or-perish (2007)
10.75 H. Etzkowitz, L. Leydesdorff: The dynamics of innovation: From national systems and "Mode 2" to a triple helix of university-industry-government relations, Res. Policy **29**, 109–123 (2000)
10.76 T.A. Brooks: Evidence of complex citer motivations, J. Am. Soc. Inf. Sci. **37**, 34–36 (1986)
10.77 M.H. MacRoberts, B.R. MacRoberts: Author motivation for not giving citing influences—a methodological note, J. Am. Soc. Inf. Sci. **39**, 432–433 (1988)
10.78 M.H. MacRoberts, B.R. MacRoberts: Problems of citation analysis, Scientometrics **36**, 435–444 (1996)
10.79 P. Vinkler: Comparative investigation of frequency and strength of motives toward referencing, the reference threshold model—comments on theories of citation?, Scientometrics **43**, 107–127 (1998)
10.80 L. Bornmann, H.D. Daniel: What do citation counts measure? A review of studies on citing behavior, J. Documentation **64**(1), 45–80 (2008)
10.81 A.F.J. van Raan: In matters of quantitative studies of science the fault of theorists is offering too little and asking too much, Scientometrics **43**, 129–139 (1998)
10.82 T.F. Frandsen, J. Nicolaisen: Citation behavior: A large-scale test of the persuasion by name-dropping hypothesis, J. Assoc. Inf. Sci. Technol. **65**(2), 1278–1284 (2017)
10.83 A.J. Nederhof, A.F.J. van Raan: Peer-review and bibliometric indicators of scientific performance—A comparison of cum laude doctorates with ordinary doctorates in physics, Scientometrics **11**(5-6), 333–350 (1987)
10.84 A.J. Nederhof, A.F.J. van Raan: A validation-study of bibliometric indicators—The comparative performance of cum laude doctorates in chemistry, Scientometrics **17**(5-6), 427–435 (1989)
10.85 D.F. Horrobin: The philosophical basis of peer review and the suppression of innovation, J. Am. Med. Assoc. **263**, 1438–1441 (1990)
10.86 C. Wennerås, A. Wold: Nepotism and sexism in peer-review, Nature **387**, 341–343 (1997)
10.87 The Wellcome Trust: *Women and Peer Review. An Audit of the Wellcome Trust Decision-Making on Grants* (The Wellcome Trust/PRISM, London 1997)
10.88 H.W. Marsh, L. Bornmann, R. Mutz, H.D. Daniel, A. O'Mara: Gender effects in the peer reviews of grant proposals: A comprehensive meta-analysis comparing traditional and multilevel approaches, Rev. Educ. Res. **79**(3), 1290–1326 (2009)
10.89 A.J. Nederhof: The validity and reliability of evaluation of scholarly performance. In: *Handbook of Quantitative Studies of Science and Technology*,

ed. by A.F.J. van Raan (Elsevier/North-Holland, Amsterdam 1988) pp. 193–228
10.90 W. Glänzel: A bibliometric approach to social sciences, national research performances in 6 selected social science areas, 1990–1992, Scientometrics **35**, 291–307 (1996)
10.91 D. Hicks: The difficulty of achieving full coverage of international social science literature and the bibliometric consequences, Scientometrics **44**, 193–215 (1999)
10.92 H.F. Moed, M. Luwel, A.J. Nederhof: Towards research performance measurement in the humanities, Libr. Trends **50**, 498–520 (2002)
10.93 L. Butler, M.S. Visser: Extending citation analysis to non-source items, Scientometrics **66**(2), 327–343 (2006)
10.94 A.J. Nederhof: Bibliometric monitoring of research performance in the social sciences and the humanities: A review, Scientometrics **66**(1), 81–100 (2006)
10.95 A.J. Nederhof, T.N. van Leeuwen, A.F.J. van Raan: Highly cited non-journal publications in political science, economics and psychology: A first exploration, Scientometrics **83**(2), 363–374 (2010)
10.96 A. Linmans: Why with bibliometrics the humanities does not need to be the weakest link. Indicators for research evaluation based on citations, library holdings, and productivity measures, Scientometrics **83**(2), 337–354 (2010)
10.97 L. Egghe, R. Rousseau: The influence of publication delays on the observed aging distribution of scientific literature, J. Am. Soc. Inf. Sci. **51**, 158–165 (2000)
10.98 E. Garfield: Premature discovery or delayed recognition –why? In: *Essays of an Information Scientist*, Vol. 4 (1980) pp. 488–493
10.99 W. Glänzel, B. Schlemmer, B. Thijs: Better late than never? On the chance to become highly cited only beyond the standard time horizon, Scientometrics **58**(3), 571–586 (2003)
10.100 H. Small, E. Greenlee: Citation context analysis of a co-citation cluster-recombinant DNA, Scientometrics **2**(4), 277–301 (1980)
10.101 H. Small, E. Sweeney: Clustering the science citation index using co-citations, I: A comparison of methods, Scientometrics **7**, 393–404 (1985)
10.102 H. Small, E. Sweeney, E. Greenlee: Clustering the science citation index using co-citations, II: Mapping science, Scientometrics **8**, 321–340 (1985)
10.103 H. Small: Visualizing science by citation mapping, J. Am. Soc. Inf. Sci. **50**, 799–813 (1999)
10.104 E. Noma: An improved method for analysing square scientometric transaction matrices, Scientometrics **4**, 297–316 (1982)
10.105 Q. Zhou, L. Leydesdorff: The normalization of occurrence and co-occurrence matrices in bibliometrics using *Cosine* similarities and *Ochiai* coefficients, J. Assoc. Inf. Sci. Technol. **67**(11), 2805–2814 (2016)
10.106 H.P.F. Peters, A.F.J. van Raan: Co-word-based science maps of chemical engineering, 1. Representations by direct multidimensional-scaling, Res. Policy **22**(1), 23–45 (1993)
10.107 H.P.F. Peters, A.F.J. van Raan: Co-word-based science maps of chemical engineering, 2. Representations by combined clustering and multidimensional-scaling, Res. Policy **22**(1), 47–71 (1993)
10.108 L. Waltman, N.J. van Eck: A new methodology for constructing a publication-level classification system of science, J. Am. Soc. Inf. Sci. Technol. **63**(12), 2378–2392 (2012)
10.109 L. Waltman, N.J. van Eck: A smart local moving algorithm for large-scale modularity-based community detection, European Physical Journal B **86**(11), 471 (2013)
10.110 N.J. van Eck, L. Waltman: CitNetExplorer: A new software tool for analyzing and visualizing citation networks, J. Informetrics **8**(4), 802–823 (2014)
10.111 K.W. Boyack, R. Klavans: Creation of a highly detailed, dynamic, global model and map of science, J. Am. Soc. Inf. Sci. Technol. **65**(4), 670–685 (2014)
10.112 CitNetExplorer: http://www.citnetexplorer.nl/home
10.113 N.J. van Eck, L. Waltman: Software survey: VOSviewer, a computer program for bibliometric mapping, Scientometrics **84**(2), 523–538 (2010)
10.114 VOSviewer: http://vosviewer.com
10.115 L. Egghe: Theory and practise of the g-index, Scientometrics **69**(1), 131–152 (2006)
10.116 K.W. Boyack, R. Klavans: Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately?, J. Am. Soc. Inf. Sci. Technol. **61**, 2389–2404 (2010)
10.117 M. Rosvall, D. Axelsson, C.T. Bergstrom: The map equation, Eur. Phys. J. ST **178**, 13–23 (2009)
10.118 M. Rosvall, C.T. Bergstrom: Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems, PLoS ONE **6**(4), e18209 (2011)
10.119 L. Subelj, N.J. van Eck, L. Waltman: Clustering scientific publications based on citation relations: A systematic comparison of different methods, PLoS ONE **11**(4), e0154404 (2016)
10.120 D.J. de Solla Price: Toward a model for science indicators. In: *Toward a Metric of Science: The Advent of Science Indicators*, ed. by Y. Elkana, J. Lederberg, R.K. Merton, A. Thackray, H. Zuckerman (John Wiley, New York 1978)
10.121 J. Ziman: From parameters to portents—and back. In: *Toward a Metric of Science: The Advent of Science Indicators*, ed. by Y. Elkana, J. Lederberg, R.K. Merton, A. Thackray, H. Zuckerman (John Wiley, New York 1978)
10.122 VSNU: *Assessment of Research Quality: Chemistry and Chemical Engineering* (VSNU, Utrecht 2002)
10.123 Uppsala University: *Quality and Renewal 2007: An Overall Evaluation of Research at Uppsala University* (Uppsala University, Uppsala 2007)
10.124 M. Olensky, M. Schmidt, N.J. van Eck: Evaluation of the citation matching algorithms of CWTS and

iFQ in comparison to the Web of Science, J. Assoc. Inf. Sci. Technol. **67**(10), 2550–2564 (2016)
10.125 L. Bornmann, L. Leydesdorff, R. Mutz: The use of percentiles and percentile rank classes in the analysis of bibliometric data: Opportunities and limits, J. Informetr. **7**(1), 158–165 (2013)
10.126 L. Waltman, N.J. van Eck: A systematic empirical comparison of different approaches for normalizing citation impact indicators, J. Informetr. **7**(4), 833–849 (2013)
10.127 L. Waltman, N.J. van Eck, T.N. van Leeuwen, M.S. Visser, A.F.J. van Raan: Towards a new crown indicator: An empirical analysis, Scientometrics **87**(3), 467–481 (2011)
10.128 L. Waltman, N.J. van Eck, T.N. van Leeuwen, M.S. Visser, A.F.J. van Raan: Towards a new crown indicator: Some theoretical considerations, J. Informetr. **5**(1), 37–47 (2011)
10.129 T. Opthof, L. Leydesdorff: Caveats for the journal and field normalizations in the CWTS ("Leiden") evaluations of research performance, J. Informetr. **4**(3), 423–430 (2010)
10.130 N.J. van Eck, L. Waltman, A.F.J. van Raan, R.J.M. Klautz, W.C. Peul: Citation analysis may severely underestimate the impact of clinical research as compared to basic research, PLoS ONE **8**(4), e62395 (2013)
10.131 Q. Wang, L. Waltman: Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus, J. Informetr. **10**, 347–364 (2016)
10.132 H.F. Moed: Measuring contextual citation impact of scientific journals, J. Informetr. **4**(3), 265–27726a (2010)
10.133 L. Leydesdorff, T. Opthof: Scopus's Source Normalized Impact per Paper (SNIP) versus a journal impact factor based on fractional counting of citations, J. Am. Soc. Inf. Sci. Technol. **61**(11), 2365–2369 (2010)
10.134 L. Leydesdorff: The revised SNIP indicator of Elsevier's Scopus, J. Informetr. **7**(4), 859–860 (2013)
10.135 L. Waltman, N.J. van Eck: Source normalized indicators of citation impact: An overview of different approaches and an empirical comparison, Scientometrics **96**(3), 699–716 (2013)
10.136 L. Waltman: A review of the literature on citation impact indicators, J. Informetr. **10**(2), 365–391 (2016)
10.137 G. Abramo, C.A. D'Angelo: A farewell to the MNCS and like size-independent indicators, J. Informetr. **10**(2), 646–651 (2016)
10.138 G. Abramo, C.A. D'Angelo: A farewell to the MNCS and like size-independent indicators: Rejoinder, J. Informetr. **10**(2), 679–683 (2016)
10.139 W. Glänzel, B. Thijs, K. Debackere: Productivity, performance, efficiency, impact—What do we measure anyway? Some comments on the paper "A farewell to the MNCS and like size-independent indicators" by Abramo and D'Angelo, J. Informetr. **10**(2), 658–660 (2016)
10.140 G. Sivertsen: A welcome to methodological pragmatism, J. Informetr. **10**(2), 664–666 (2016)
10.141 L. Bornmann, R. Haunschild: Efficiency of research performance and the glass researcher, J. Informetr. **10**(2), 652–654 (2016)
10.142 R. Danell: Evaluating research organizations' contribution to science is not the same task as evaluating the performance of their scientists, J. Informetr. **10**(2), 655–657 (2016)
10.143 M. Thelwall: Not dead, just resting: The practical value of per publication citation indicators, J. Informetr. **10**(2), 667–670 (2016)
10.144 M. Zitt: Paving the way or pushing at open doors? A comment on Abramo and D'Angelo "Farewell to size-independent indicators", J. Informetr. **10**(2), 675–678 (2016)
10.145 J. Ruiz-Castillo: Research output indicators are not productivity indicators, J. Informetr. **10**(2), 661–663 (2016)
10.146 L. Waltman, N.J. van Eck, M. Visser, P. Wouters: The elephant in the room: The problem of quantifying productivity in evaluative scientometrics, J. Informetr. **10**(2), 671–674 (2016)
10.147 A.F.J. van Raan: Performance-related differences of bibliometric statistical properties of research groups: cumulative advantages and hierarchically layered networks, J. Am. Soc. Inf. Sci. Technol. **57**(14), 1919–1935 (2006)
10.148 A.F.J. van Raan: Scaling rules in the science system: Influence of field-specific citation characteristics on the impact of research groups, J. Am. Soc. Inf. Sci. Technol. **59**(4), 565–576 (2008)
10.149 L. Leydesdorff, T. Opthof: Normalization at the field level: Fractional counting of citations, J. Informetr. **4**(4), 644–646 (2010)
10.150 L. Leydesdorff, L. Bornmann: How fractional counting of citations affects the impact factor: Normalization in terms of differences in citation potentials among fields of science, J. Am. Soc. Inf. Sci. Technol. **62**(2), 217–229 (2011)
10.151 J. Ruiz-Castillo, L. Waltman: Field-normalized citation impact indicators using algorithmically constructed classification systems of science, J. Informetr. **9**(1), 102–117 (2015)
10.152 L. Waltman, N.J. van Eck: Field-normalized citation impact indicators and the choice of an appropriate counting method, J. Informetr. **9**(4), 872–894 (2015)
10.153 R. van Noorden: Love thy lab neighbour, Nature **468**, 1011 (2010)
10.154 K. Lee, J.S. Brownstein, R.G. Mills, I.S. Kohane: Does collocation inform the impact of collaboration?, PLoS ONE **5**(12), e14279 (2010)
10.155 A. Gazni, C.R. Sugimoto, F. Didegah: Mapping world scientific collaboration: Authors, institutions, and countries, J. Am. Soc. Inf. Sci. Technol. **63**(2), 323–335 (2012)
10.156 P. Zhou, R.J.W. Tijssen, L. Leydesdorff: University-industry collaboration in China and the USA: A bibliometric comparison, PLoS ONE **11**(11), e0165277 (2016)
10.157 A.F.J. van Raan: On growth, ageing, and fractal differentiation of science, Scientometrics **47**(2), 347–362 (2000)

10.158 D. Hicks, P. Wouters, L. Waltman, S. de Rijcke, I. Rafols: The Leiden Manifesto for research metrics, Nature **520**(7548), 429–431 (2015)
10.159 A.F.J. van Raan: Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises, Scientometrics **36**, 397–420 (1996)
10.160 A.F.J. van Raan: Scientometrics: State-of-the-Art, Scientometrics **38**, 205–218 (1997)
10.161 Wikipedia: Retraction Watch, https://en.wikipedia.org/wiki/Retraction_Watch
10.162 J.A. Byrne, C. Labbé: Striking similarities between publications from China describing single gene knockdown experiments in human cancer cell lines, Scientometrics **110**(3), 1471–1493 (2017)
10.163 T.N. van Leeuwen, M. Luwel: An in-depth analysis of papers retracted in the Web of Science. In: *Proc. 19th Int. Conf. Sci. Technol. Indic., Leiden* (2014) pp. 337–344
10.164 A.F.J. van Raan, T.N. van Leeuwen, M.S. Visser: Severe language effect in university rankings: particularly Germany and France are wronged in citation-based rankings, Scientometrics **88**(2), 495–498 (2011)
10.165 A.F.J. van Raan, T.N. van Leeuwen, M.S. Visser: Non-English papers decrease rankings, Nature **469**(7328), 34 (2011)
10.166 A.F.J. van Raan: Sleeping beauties in science, Scientometrics **59**(3), 461–466 (2004)
10.167 T. Braun, W. Glänzel, A. Schubert: On sleeping beauties, princes and other tales of citation distributions, Res. Eval. **19**(3), 195–202 (2010)
10.168 C. Lachance, V. Larivière: On the citation lifecycle of papers with delayed recognition, J. Informetr. **8**, 863–872 (2014)
10.169 A.F.J. van Raan: Dormitory of physical and engineering sciences: Sleeping beauties may be sleeping innovations, PLoS ONE **10**(10), e0139786 (2015)
10.170 P. Gorry, P. Ragouet: "Sleeping beauty" and her restless sleep: Charles Dotter and the birth of interventional radiology, Scientometrics **107**(2), 773–784 (2016)
10.171 A.F.J. van Raan: Sleeping beauties cited in patents: Is there also a dormitory of inventions?, Scientometrics **110**(3), 1123–1156 (2017)
10.172 Q.L. Burrell: Are "sleeping beauties" to be expected?, Scientometrics **65**(3), 381–389 (2005)
10.173 J. Wang, F. Ma, M. Chen, Y. Rao: Why and how can sleeping beauties be awakened?, Electron. Libr. **30**(1), 5–18 (2012)
10.174 B. van Calster: It takes time: a remarkable example of delayed recognition, J. Am. Soc. Inf. Sci. Technol. **63**(11), 2341–2344 (2012)
10.175 J. Li, F.Y. Ye: The phenomenon of all-elements-sleeping-beauties in scientific literature, Scientometrics **92**(3), 795–799 (2012)
10.176 J. Li, D. Shi, S.X. Zhao, F.Y. Ye: A study of the "heartbeat spectra" for "sleeping beauties", J. Informetr. **8**, 493–502 (2014)
10.177 J. Li: Citation curves of "all-elements-sleeping-beauties": "flash in the pan" first and then "delayed recognition", Scientometrics **100**(2), 595–601 (2014)
10.178 G.A. Ronda-Pupo, J.S. Katz: The scaling relationship between citation-based performance and coauthorship patterns in natural sciences, J. Assoc. Inf. Sci. Technol. **68**(5), 1257–1265 (2017)
10.179 T.N. van Leeuwen: Discussing some basic critique on journal impact factors: Revision of earlier comments, Scientometrics **92**(2), 443–455 (2012)
10.180 A.F.J. van Raan: Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups, Scientometrics **67**(3), 491–502 (2006)
10.181 L. Bornmann, R. Mutz, H.D. Daniel: Are there better indices for evaluation purposes than the h-index? A comparison of nine different variants of the h-index using data from biomedicine, J. Am. Soc. Inf. Sci. Technol. **59**(5), 830–837 (2008)
10.182 L. Waltman, N.J. Van Eck: The inconsistency of the h-index, J. Am. Soc. Inf. Sci. Technol. **63**(2), 406–415 (2012)
10.183 E.J. Rinia, T.N. Van Leeuwen, H.G. van Vuren, A.F.J. van Raan: Comparative analysis of a set of bibliometric indicators and central peer review criteria. Evaluation of condensed matter physics in the Netherlands, Res. Policy **27**, 95–107 (1998)
10.184 E.J. Rinia, T.N. van Leeuwen, H.G. van Vuren, A.F.J. van Raan: Influence of interdisciplinarity on peer-review and bibliometric evaluations, Res. Policy **30**, 357–361 (2001)
10.185 H.F. Moed: *Citation Analysis in Research Evaluation* (Springer, Dordrecht 2005) pp. 229–257
10.186 L. Bornmann, H.D. Daniel: Selecting scientific excellence through committee peer review—A citation analysis of publications previously published to approval or rejection of post-doctoral research fellowship applicants, Scientometrics **68**(3), 427–440 (2006)
10.187 Y. Gargouri, C. Hajjem, V. Larivière, Y. Gingras, L. Carr, T. Brody, S. Harnad: Self-selected or mandated, open access increases citation impact for higher quality research, PLoS ONE **5**(10), e13636 (2010)
10.188 Dimensions database: https://app.dimensions.ai/discover/publication
10.189 H.F. Moed: The effect of "Open access" on citation impact: An analysis of ArXiv's condensed matter section, J. Am. Soc. Inf. Sci. Technol. **58**(13), 2047–2054 (2007)
10.190 A. Kastberg: Is open access really good for science?, Europhys. News **45**(2), 32 (2014)
10.191 L. Björneborn, P. Ingwersen: Perspectives of webometrics, Scientometrics **50**, 65–82 (2001)
10.192 M. Thelwall, A. Smith: Interlinking between Asia-Pacific University Web sites, Scientometrics **55**, 363–376 (2002)
10.193 M. Thelwall, G. Harries: The connection between the research of a university and counts of links to its web pages: An investigation based upon a classification of the relationships of pages to the research of the host university, J. Am. Soc. Inf. Sci. **54**, 594–602 (2003)

10.194 I.F. Aguillo: Building web indicators for the EU OA repository. In: *Workshop on New Research Lines in Informetrics, IPP-CCHS (CSIC), Madrid* (2011), http://digital.csic.es/bitstream/10261/40279/1/OpenAIRE%20Webometrics.pdf
10.195 J. Priem, D. Taraborelli, P. Groth, C. Neylon: Altmetrics: A manifesto, http://altmetrics.org/manifesto/ (2010)
10.196 J. Priem, P. Groth, D. Taraborelli: The altmetrics collection, PLoS ONE **7**(11), e48753 (2012)
10.197 J. Bar-Ilan: Data collection methods on the web for informetric purposes—A review and analysis, Scientometrics **50**, 7–32 (2001)
10.198 C. Schlögl, J. Gorraiz: Global usage versus global citation metrics: The case of pharmacology journals, J. Am. Soc. Inf. Sci. Technol. **62**(1), 161–170 (2011)
10.199 C. Schlögl, J. Gorraiz, C. Gumpenberger, K. Jack, P. Kraker: Comparison of downloads, citations and readership data for two information systems journals, Scientometrics **101**(2), 1113–1128 (2014)
10.200 J. Bar-Ilan, S. Haustein, I. Peters, J. Priem, H. Shema, J. Terliesner: Beyond citations: Scholars' visibility on the social web. In: *Proc. 17th Int. Conf. Sci. Technol. Indic., Montréal* (2012) pp. 98–109
10.201 H. Shema, J. Bar-Ilan, M. Thelwall: Do blog citations correlate with a higher number of future citations? Research blogs as a potential source for alternative metrics, J. Assoc. Inf. Sci. Technol. **65**(5), 1018–1027 (2014)
10.202 M. Thelwall, S. Haustein, V. Larivière, C. Sugimoto: Do altmetrics work? Twitter and ten other candidates, PLoS ONE **8**(5), e64841 (2013)
10.203 Research Trends: Special issue on alternative metrics, Issue 37, June (2014)
10.204 R. Costas, Z. Zahedi, P. Wouters: Do altmetrics correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective, J. Assoc. Inf. Sci. Technol. **66**(10), 2003–2019 (2014)
10.205 P. Sud, M. Thelwall: Evaluating altmetrics, Scientometrics **98**(2), 1131–1143 (2014)
10.206 C. Chorus, L. Waltman: A large-scale analysis of impact factor biased journal self-citations, PLoS ONE **11**(8), e0161021 (2016)
10.207 Wikipedia: San Francisco Declaration on Research Assessment, https://en.wikipedia.org/wiki/San_Francisco_Declaration_on_Research_Assessment
10.208 L. Waltman, V.A. Traag: Use of the journal impact factor for assessing individual articles need not be wrong, https://arxiv.org/abs/1703.02334 (2017)
10.209 M. Zitt, H. Small: Modifying the journal impact factor by fractional citation weighting: The audience factor, J. Am. Soc. Inf. Sci. Technol. **59**(11), 1856–1860 (2008)
10.210 L. Waltman, N.J. van Eck: The relation between Eigenfactor, audience factor, and influence weight, J. Am. Soc. Inf. Sci. Technol. **61**(7), 1476–1486 (2010)
10.211 CWTS Journal Indicators: http://www.journalindicators.com/
10.212 V.P. Guerrero-Botea, F. Moya-Anegón: A further step forward in measuring journals' scientific prestige: The SJR2 indicator, J. Informetr. **6**, 674–688 (2012)
10.213 Scimago journal & country rank: http://www.scimagojr.com/journalrank.php
10.214 Elsevier: CiteScore, https://www.elsevier.com/editors-update/story/journal-metrics/citescore-a-new-metric-to-help-you-choose-the-right-journal
10.215 F. Franceschini, D. Maisano, L. Mastrogiacomo: The museum of errors/horrors in Scopus, J. Informetr. **10**(1), 174–182 (2016)
10.216 W.J.N. Meester, L. Colledge, E.E. Dyas: A response to "The museum of errors/horrors in Scopus" by Franceschini et al, J. Informetr. **10**(1), 569–570 (2016)
10.217 S. Brin, L. Page: The anatomy of a large-scale hypertextual web search engine, Comput. Netw. ISDN Syst. **30**(1–7), 107–117 (1998)
10.218 C.T. Bergstrom: Eigenfactor: Measuring the value and prestige of scholarly journals, Coll. Res. Libr. News **68**(5), 314–316 (2007)
10.219 B. Zogala-Siudem, G. Siudem, A. Cena, M. Gagolewski: Agent-based model for the h-index—exact solution, Eur. Phys. J. B **89**, 21 (2016)
10.220 Research Policy Team HEFCE: REF Independent review of the role of metrics in research assessment. Summary of responses submitted to the call for evidence, http://dera.ioe.ac.uk/21370/2/Summary%20of%20responses%20submitted%20to%20the%20call%20for%20evidence.pdf (Oct. 2014)
10.221 H.F. Moed, J. Bar-Ilan, G. Halevi: A new methodology for comparing Google Scholar and Scopus, J. Informetr. **10**(2), 533–551 (2016)
10.222 R. van. Noorden: Science research engine links papers to grants and patents. The Dimensions database promises a financial perspective on scholarly literature, https://www.nature.com/articles/d41586-018-00688-0 (16 Jan. 2018)
10.223 D.M. Nichols, M.B. Twidale: Metrics for openness, J. Assoc. Inf. Sci. Technol. **68**(4), 1048–1060 (2016)
10.224 Webometrics Ranking of World Universities: http://www.webometrics.info/en
10.225 M. Thelwall, K. Kousha: ResearchGate articles: Age, discipline, audience size, and impact, J. Assoc. Inf. Sci. Technol. **68**(2), 468–479 (2017)
10.226 PLoS ONE: Altmetrics, http://journals.plos.org/plosone/browse/altmetrics
10.227 Oxford University Press: https://academic.oup.com/journals/pages/authors/article_level_metrics
10.228 Altmetrics: http://altmetrics.org
10.229 Impactstory: https://profiles.impactstory.org/
10.230 R. Haunschild, L. Bornmann: Normalization of Mendeley reader counts for impact assessment, J. Informetr. **10**(1), 62–73 (2016)
10.231 M. Thelwall, P. Sud: Mendeley readership counts: An investigation of temporal and disciplinary dif-

ferences, J. Assoc. Inf. Sci. Technol. **67**(12), 3036–3050 (2016)
10.232 National Information Standards Organization (NISO): NISO Alternative Assessment Metrics (Altmetrics) Initiative, https://www.niso.org/standards-committees/altmetrics
10.233 The Scientist: Top 10 Retractions of 2016, http://www.the-scientist.com/?articles.view/articleNo/47813/title/Top-10-Retractions-of-2016/
10.234 H.F. Moed: A critical comparative analysis of five world university rankings, Scientometrics **110**(2), 967–990 (2017)
10.235 M. Dobrota, M. Bulajic, L. Bornmann, V. Jeremic: A new approach to the QS university ranking using the composite i-distance indicator: Uncertainty and sensitivity analyses, J. Assoc. Inf. Sci. Technol. **67**(1), 200–211 (2016)
10.236 A. Perianes-Rodriguez, J. Ruiz-Castillo: University citation distributions, J. Assoc. Inf. Sci. Technol. **67**(11), 2790–2804 (2016)
10.237 C. Daraio, A. Bonaccorsi: Beyond university rankings? Generating new indicators on universities by linking data in open platforms, J. Assoc. Inf. Sci. Technol. **68**(2), 508–529 (2017)
10.238 Committee on Publication Ethics (COPE): Promoting integrity in research and its publication, http://publicationethics.org/
10.239 E. Delgado Lopez-Cozar, N. Robinson-Garcia, D. Torres-Salinas: The Google Scholar experiment: How to index false papers and manipulate bibliometric indicators, J. Assoc. Inf. Sci. Technol. **65**(3), 446–454 (2014)
10.240 L. Butler: Modifying publication practices in response to funding formulas, Res. Eval. **17**, 39–46 (2003)
10.241 S. De Rijcke, P.F. Wouters, A.D. Rushforth, T.P. Franssen, B. Hammarfelt: Evaluation practices and effects of indicator use—A literature review, Res. Eval. **25**(2), 161–169 (2016)
10.242 V. Larivière, R. Costas: How many is too many? On the relationship between research productivity and impact, PLoS ONE **11**(9), e0162709 (2016)
10.243 N. Caplar, S. Tacchella, S. Birrer: Quantitative evaluation of gender bias in astronomical publications from citation counts, Nat. Astron. **1**(6), 0141 (2017)
10.244 J. Wilsdon, L. Allen, E. Belfiore, P. Campbell, S. Curry, S. Hill, R. Jones, R. Kain, S. Kerridge, M. Thelwall, J. Tinkler, I. Viney, P. Wouters, J. Hill, B. Johnson: *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management* (HEFCE, London 2015)
10.245 A. Ancaiani, A.F. Anfossi, A. Barbara, S. Benedetto, B. Blasi, V. Carletti, T. Cicero, A. Ciolfi, F. Costa, G. Colizza, M. Costantini, F. di Cristina, A. Ferrara, R.M. Lacatena, M. Malgarini, I. Mazzotta, C.A. Nappi, S. Romagnosi, S. Sileoni: Evaluating scientific research in Italy: The 2004–10 research evaluation exercise, Res. Eval. **24**, 242–255 (2015)
10.246 A. Molinié, G. Bodenhausen: Bibliometrics as weapons of mass citation, Chimia **64**, 78–89 (2010)
10.247 R.R. Ernst: The follies of citation indices and academic ranking lists. A brief commentray to 'Bibliometrics as weapons of mass citation', Chimia **64**, 90 (2010)
10.248 Wikipedia: Ludolph van Ceulen, https://en.wikipedia.org/wiki/Ludolph_van_Ceulen
10.249 E. Noyons, C. Calero Medina: *Results of the bibliometric study on the Faculty of Veterinary Medicine of the Utrecht University 2001–2011*, CWTS Report April 2012, Center for Science Technol. Studies (CWTS) (Leiden University, Leiden 2012), https://www.uu.nl/en/organisation/faculty-of-veterinary-medicine/veterinary-research/organisation/assessment-of-research-quality
10.250 R. Costas, T.N. van Leeuwen, A.F.J. van Raan: Is scientific literature subject to a 'sell-by-date'? A general methodology to analyze the 'durability' of scientific documents, J. Am. Soc. Inf. Sci. Technol. **61**(2), 329–339 (2010)
10.251 R. Costas, T.N. van Leeuwen, A.F.J. van Raan: The "Mendel syndrome" in science: Durability of scientific literature and its effects on bibliometric analysis of individual scientists, Scientometrics **89**(1), 177–205 (2011)
10.252 R. Costas, T.N. van Leeuwen, A.F.J. van Raan: Effects of the durability of scientific literature at the group level: Case study of chemistry research groups in the Netherlands, Res. Policy **42**(4), 886–894 (2013)
10.253 A.F.J. van Raan, T.N. van Leeuwen: Assessment of the scientific basis of interdisciplinary, applied research: Application of bibliometric methods in nutrition and food research, Res. Policy **31**, 611–632 (2002)
10.254 CWTS B.V.: Monitoring & Evaluation, Leiden University, http://www.cwtsbv.nl/Monitoring-Evaluation
10.255 M. Thelwall, P. Sud: National, disciplinary and temporal variations in the extent to which articles with more authors have more impact: Evidence from a geometric field normalised citation indicator, J. Informetr. **10**(1), 48–61 (2016)
10.256 W. Glänzel, M. Meyer: Patents cited in the scientific literature: An exploratory study of 'reverse' citation relations, Scientometrics **58**, 415–428 (2003)
10.257 A.F.J. van Raan: Patent citations analysis and its value in research evaluation: A review and a new approach to map technology-relevant research, J. Data Inf. Sci. **2**(1), 13–50 (2017)
10.258 B. Uzzi, S. Mukherjee, M. Stringer, B. Jones: Atypical combinations and scientific impact, Science **342**, 468–471 (2013)
10.259 A. Thor, W. Marx, L. Leydesdorff, L. Bornmann: Introducing CitedReferencesExplorer (CRExplorer): A program for reference publication year spectroscopy with cited references standardization, J. Informetr. **10**(2), 503–515 (2016)

10.260 G. Chen, L. Xiao: Selecting publication keywords for domain analysis in bibliometrics: A comparison of three methods, J. Informetr. **10**(1), 212–213 (2016)
10.261 K.H. Lee, S.Y. Kim, H.J. Kim, M. Song: Comparative evaluation of bibliometric content networks by tomographic content analysis: An application to Parkinson's disease, J. Assoc. Inf. Sci. Technol. (2017), https://doi.org/10.1002/asi.23752
10.262 L. Leydesdorff, A. Nerghes: Co-word maps and topic modeling: A comparison using small and medium-sized corpora ($N < 1000$), J. Assoc. Inf. Sci. Technol. **68**(4), 1024–1035 (2017)
10.263 A. Suominen, H. Toivanen: Map of science with topic modeling: Comparison of unsupervised learning and human-assigned subject classification, J. Assoc. Inf. Sci. Technol. **67**(10), 2464–2476 (2016)
10.264 B. Wen, E. Horlings, M. van der Zouwen, P. van den Besselaar: Mapping science through bibliometric triangulation: An experimental approach applied to water research, J. Assoc. Inf. Sci. Technol. **68**(3), 724–738 (2017)
10.265 E.C.M. Noyons, A.F.J. van Raan: Monitoring scientific developments from a dynamic perspective: Self-organized structuring to map neural network research, J. Am. Soc. Inf. Sci. **49**(1), 68–81 (1998)
10.266 L. Waltman, A.F.J. van Raan, S. Smart: Exploring the relationship between the engineering and physical sciences and the health and life sciences by advanced bibliometric methods, PLoS ONE **9**(10), e111530 (2014)
10.267 L. Leydesdorff, I. Rafols: A global map of science based on the ISI subject categories, J. Am. Soc. Inf. Sci. Technol. **60**(2), 348–362 (2009)
10.268 L. Leydesdorff, S. Carley, I. Rafols: Global maps of science based on the new Web-of-Science categories, Scientometrics **94**(2), 589–593 (2013)
10.269 H. Small, K.W. Boyack, R. Klavans: Identifying emerging topics in science and technology, Res. Policy **43**, 1450–1467 (2014)
10.270 D. Zhao, A. Strotmann: Can citation analysis of web publications better detect research fronts?, J. Am. Soc. Inf. Sci. Technol. **58**(9), 1285–1302 (2007)
10.271 S. Jensen, X. Liub, Y. Yuc, S. Milojevic: Generation of topic evolution trees from heterogeneous bibliographic networks, J. Informetr. **10**, 606–621 (2016)
10.272 G. Holton: Can science be measured? In: *Toward a metric of science: the advent of science indicators*, ed. by Y. Elkana, J. Lederberg, R.K. Merton, A. Thackray, H. Zuckerman (John Wiley, New York 1978)

Anthony F.J. van Raan

Centre for Science and Technology Studies (CWTS)
Leiden University
Leiden, The Netherlands
vanraan@cwts.leidenuniv.nl

Dr. Anthony (Ton) F.J. van Raan is Professor emeritus of Quantitative Studies of Science at Leiden University. He holds a PhD in physics (Utrecht, 1973). His main interests involve the application of bibliometric indicators in research evaluation, mapping of science, statistical properties of indicators, and the scaling behavior of universities and cities. He has published (as author and co-author) around 30 articles in physics and 200 in science and technology studies.

11. Field Normalization of Scientometric Indicators

Ludo Waltman, Nees Jan van Eck

When scientometric indicators are used to compare research units active in different scientific fields, there is often a need to make corrections for differences between fields, for instance, differences in publication, collaboration, and citation practices. Field-normalized indicators aim to make such corrections. The design of these indicators is a significant challenge. We discuss the main issues in the design of field-normalized indicators and present an overview of the different approaches that have been developed for dealing with the problem of field normalization. We also discuss how field-normalized indicators can be evaluated and consider the sensitivity of scientometric analyses to the choice of a field-normalization approach.

11.1 Background

Many scientometric analyses are restricted to a single field of science, but scientometric analyses also commonly stretch out over multiple scientific fields, and they often even aim to cover the entire scientific universe. University rankings, for instance, rely on scientometric indicators that are supposed to provide meaningful information about the performance of universities across many different fields of science. Likewise, many universities regularly carry out scientometric analyses in which they compare their performance in different scientific fields.

Scientific fields, of course, differ from each other in many ways, and some of these differences have important implications for scientometric analyses. For instance, in some fields, researchers tend to produce many more outputs than in other fields. In some fields, researchers focus on publishing journal articles, while in other fields they are more interested in writing books. In some fields, researchers work together in large collaborative teams, often resulting in publications with many co-authors, while in other fields researchers prefer to work individually or in small teams. In some fields, researchers cite a lot, while in other fields they cite much more sparingly, and in some fields they mainly cite recent work, while in other fields they prefer to cite older literature.

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_11

Given these differences between scientific fields, it is clear that the interpretation of a scientometric analysis that covers multiple fields is far from straightforward. Suppose that a biologist has produced 25 publications during the past 5 years, while an economist has produced 10 publications during the same time period. Can it be concluded that the biologist has been more productive than the economist? This depends on our understanding of the concept of productivity. If productivity is understood simply as the number of publications produced during a certain time period, the biologist has obviously been more productive than the economist. However, in many cases, we are probably interested in a more sophisticated concept of productivity. We may have in mind a concept of productivity that accounts for differences between fields in the rate at which researchers tend to produce publications. Based on such a more refined notion of productivity, the answer to our question is much less obvious. It may actually turn out that from this perspective, the economist should be considered more productive than the biologist.

To capture the more sophisticated concept of productivity suggested above, we need a scientometric indicator that in some way corrects for differences between scientific fields in the typical number of publications produced by a researcher. Such an indicator is referred to as a field-normalized indicator. Field-normalized indicators can be constructed not only for the concept of productivity, but also for other scientometric concepts. In the literature, field-normalized indicators of scientific impact, calculated based on citation counts, have received the most attention, and they will also play a prominent role in this chapter.

The design of field-normalized indicators is a significant challenge. In this chapter, we discuss the main issues in the design of these indicators. We present an overview of the different approaches that have been developed for dealing with the problem of field normalization. We also discuss how field-normalized indicators can be evaluated and consider the sensitivity of scientometric analyses to the choice of a field-normalization approach.

This chapter partly builds on a recent review of the literature on citation impact indicators published by one of the authors [11.1].

11.2 What Is Field Normalization?

It is notoriously difficult to define in a precise way what is meant by field normalization. A precise definition of the idea of field normalization requires a definition of the notion of field. It also requires a clear perspective on the way in which scientometric indicators are affected by differences between fields. As we will explain below, these requirements are challenging, and, therefore, the idea of field normalization will almost inevitably remain somewhat ill defined.

Defining the notion of a field is far from straightforward. There is a lack of standardized terminology. No agreement exists on the differences between the term *field* and terms such as *area*, *discipline*, *domain*, *specialty*, and *topic*. In fact, these terms often seem to be used more or less interchangeably. More fundamentally, the idea of a field can be conceptualized in different ways. A useful overview of different conceptualizations is provided by *Sugimoto* and *Weingart* [11.2]. They distinguish between cognitive, social, communicative, and institutional perspectives as well as perspectives based on separatedness and tradition. Each of these perspectives provides a different understanding of the idea of a field.

Defining the notion of a field is made even more difficult by the fact that science is structured in a hierarchical way, allowing fields to be identified at different hierarchical levels [11.3]. For instance, depending on the hierarchical level that one prefers, citation analysis, bibliometrics, information science, and social sciences could all be seen as fields. Moreover, even when one focuses as much as possible on a single hierarchical level, fields typically will not be neatly separated from each other. For instance, bibliometrics, scientometrics, and research evaluation could perhaps be regarded as fields at more or less the same hierarchical level. However, it is clear that these fields are strongly interrelated and have a considerable overlap.

Field normalization of scientometric indicators is motivated by the idea that differences between fields lead to distortions in scientometric indicators. One could think of this in terms of signal and noise. Scientometric indicators provide a signal of concepts such as productivity or scientific impact, but they are also affected by noise. This noise may partly be due to differences between fields, for instance, differences in publication, collaboration, and citation practices. Field normalization aims to remove this noise while maintaining the signal.

However, the distinction between signal and noise is much less clear than it may seem at first sight. To illustrate this, let us consider citation-based indicators of scientific impact. Publications in information science on average are cited much less frequently than publications in, for instance, the life sciences. A citation-based indicator that does not account for this may be considered very noisy. The indicator may be seen as strongly biased against information science research. Suppose, therefore, that we use an indicator that corrects for differences in *citation density* between information science and other fields. Let us now zoom in on information science. Within information science, publications in scientometrics on average receive significantly more citations than publications in, for instance, library science. Again, we may feel that our indicator is too noisy and that we need to get rid of the noise. Consequently, suppose that we use an indicator that corrects for differences in citation density not only between information science and other fields, but also between scientometrics and other subfields within information science. We now zoom in on scientometrics. Within scientometrics, publications on citation analysis tend to receive more citations than publications on a topic such as co-authorship analysis. This may also be seen as noise that we need to get rid of. The next step then may be to use an indicator that corrects for differences in citation density even between different topics within scientometrics. However, we could, of course, argue that even this indicator is noisy. Suppose that empirical publications on citation analysis are cited more frequently than theoretical publications. This could then be claimed to show that the empirical publications have a higher impact than the theoretical publications. On the other hand, we could also argue that empirical and theoretical research on citation analysis represent two different subtopics and that we need to get rid of noise due to differences in citation density between these subtopics. However, if we keep following such a reasoning, at some point everything is considered noise, and there is no signal left, meaning that indicators become completely non-informative.

The above example illustrates that there is no objective way of distinguishing between signal and noise. We may say that scientometric indicators are distorted by noise that is due to differences between fields. However, fields can reasonably be defined at different hierarchical levels, leading to different perspectives on what should count as a signal and what should be seen as noise. When working with field-normalized indicators, choosing a certain hierarchical level for defining fields, and consequently making a certain distinction between signal and noise, is a normative decision. There is no objective way in which this choice can be made. Probably there is agreement that fields should not be defined in a very broad or very narrow way, but this still leaves open many intermediate ways in which fields can be defined. A single optimal way of defining fields does not exist [11.3]. Ideally, the hierarchical level at which fields are defined is chosen in such a way that it aligns well with the purpose of a specific scientometric analysis. In some analyses, it may be desirable to work with relatively narrow fields, while in other analyses broader fields may be appropriate.

11.3 Field Classification Systems

Most field-normalized indicators require an operationalization of scientific fields. We refer to such an operationalization as a field classification system. Different types of field classification systems can be distinguished. We make a distinction between classification systems of journals, publications, and researchers. Many different classification systems exist. We do not aim to provide a comprehensive overview of these systems. Instead, we focus specifically on classification systems that have been used for field-normalization purposes, either in the scientometric literature or in applied scientometric work. Each of the classification systems discussed below deals in a different way with the challenges in operationalizing scientific fields.

11.3.1 Field Classification Systems of Journals

The field classification systems used most frequently by field-normalized indicators are journal-based systems. In these systems, each journal is assigned to one or more fields. Some journal-based classification systems do not allow fields to overlap. A journal can be assigned to only one field in such systems. However, in most journal-based classification systems, overlap of fields is allowed, in which case a journal may belong to multiple fields. Some journal-based classification systems have a hierarchical structure and consist of multiple levels. Each field at a lower level is then considered to be part of a field at a higher level.

The Web of Science (WoS) database offers a classification system in which each journal indexed in the database is assigned to one or more fields. These fields are referred to as categories. There are about 250 fields in the WoS classification system.

A somewhat similar classification system is made available in the Scopus database. This system is referred to as the All Science Journal Classification (ASJC). The system has a hierarchical structure consisting of two levels. There are over 300 fields at the bottom level.

These fields have been aggregated into 27 fields at the top level. Each journal indexed in Scopus belongs to one or more fields. A comparison of the accuracy of the WoS and Scopus classification systems is reported in a study by *Wang* and *Waltman* [11.4]. According to this study, the WoS classification system is significantly more accurate than the Scopus classification system.

In the Essential Science Indicators, a tool that is based on the WoS database, a classification system of 22 broad fields is made available. In this system, it is not possible for a journal to belong to multiple fields. Each journal is assigned exclusively to a single field.

Other journal-based classification systems include the classification system of the US National Science Foundation, the classification system developed by Science-Metrix, and the classification system of *Glänzel* and *Schubert* [11.5]. The classification system of the National Science Foundation covers 125 fields, which have been aggregated into 13 broad fields. A journal can belong to only one field in this system. The system has been used in the Science and Engineering Indicators reports prepared by the National Science Foundation for a long time. Science-Metrix is a company specialized in research evaluation that has developed its own classification system. This system has been made freely available. It includes 176 fields, aggregated into 22 broad fields, with each journal being assigned to only one field. We refer to *Archambault* et al. [11.6] for more details on the approach that was taken to construct the Science-Metrix classification system. The classification system of *Glänzel* and *Schubert* [11.5] consists of two levels. The 67 fields at the bottom level have been aggregated into 15 fields at the top level. Journals may be assigned to more than one field in this system.

Multidisciplinary journals with a broad scope represent a significant challenge for journal-based classification systems. *Nature*, *Proceedings of the National Academy of Sciences*, and *Science* are well-known examples of such journals. Other examples are open access *mega journals* such as *PLOS ONE* and *Scientific Reports*. In a journal-based classification system, these multidisciplinary journals are typically assigned to a special category. In the WoS categories classification system, this category is, for instance, called *Multidisciplinary Sciences*. In the Scopus ASJC classification system, it is referred to as *multidisciplinary*. The use of a special category for multidisciplinary journals is problematic because such a category clearly does not represent a scientific field. In practice, this problem is often addressed by creating a publication-based classification system for publications in multidisciplinary journals and by complementing a journal-based classification system with such a publication-based classification system. This approach, introduced by *Glänzel* et al. [11.7]; see also [11.5], has been widely adopted. Of course, there is always some arbitrariness in deciding which journals should be considered multidisciplinary. It is clear that journals such as the ones mentioned above are of a multidisciplinary nature. However, it may be argued that journals such as *The Lancet*, *New England Journal of Medicine*, and *Physical Review Letters* should also be considered multidisciplinary and that it would be preferable to create a publication-based classification system for publications in these journals.

11.3.2 Field Classification Systems of Publications

Instead of journal-based field classification systems, it is also possible to use publication-based field classification systems. Publication-based classification systems potentially offer a more accurate and more fine-grained representation of scientific fields than their journal-based counterparts. Most publication-based classification systems are restricted to a single scientific discipline. Algorithmically constructed classification systems are an exception and may cover all scientific fields.

There are various scientific disciplines that have their own publication-based classification system. These systems often have a hierarchical structure, and they usually allow publications to be assigned to multiple fields. The use of these systems in field-normalized indicators was studied by *Bornmann* et al. [11.8], *Neuhaus* and *Daniel* [11.9], *Radicchi* and *Castellano* [11.10], and *van Leeuwen* and *Calero Medina* [11.11]. These authors focused on, respectively, the Medical Subject Headings, the Chemical Abstracts sections, the Physics and Astronomy Classification Scheme, and the EconLit classification system. Like in the case of the journal-based classification systems discussed in Sect. 11.3.1, it is important to be aware that publication-based classification systems such as the ones mentioned above were not designed specifically for field-normalization purposes.

Publication-based classification systems that are constructed algorithmically may cover all scientific fields rather than only fields within a single discipline. An approach for the algorithmic construction of publication-based classification systems was proposed by *Waltman* and *van Eck* [11.12]. In this approach, a classification system is constructed by clustering publications based on direct citation relations. Each publication is assigned to only one field. The use of algorithmically constructed publication-based classification systems in field-normalized indicators was studied by *Ruiz-Castillo* and *Waltman* [11.13] and *Perianes-Rodriguez* and *Ruiz-Castillo* [11.14]. A practical application can be found in the CWTS Leiden Ranking, a bibliometric ranking of major universities worldwide that is available at http://www.leidenranking.com. In this ranking, citation-based indicators of scientific impact are normalized using an algorithmically constructed publication-based classification system in which about 4000 scientific fields are distinguished.

An algorithmic approach to the construction of a publication-based classification system is also taken in Microsoft Academic, a recently introduced bibliometric data source somewhat similar to Google Scholar. *Hug* et al. [11.15] found that fields in the classification system of Microsoft Academic are too specific and not coherent, leading them to conclude that the classification system of Microsoft Academic is not suitable for field-normalization purposes.

11.3.3 Field Classification Systems of Researchers

Field classification systems of researchers represent a quite different approach to operationalize scientific fields. The use of researcher-based classification systems in field-normalized indicators is much less common than the use of journal-based and publication-based classification systems. Below we discuss two researcher-based classification systems that have been used in the scientometric literature.

Giovanni Abramo and *Ciriaco Andrea D'Angelo* have published a large number of papers in which they use the official classification system of Italian researchers [11.16]. This is a hierarchical system consisting of two levels. At the top level, 14 fields are distinguished. These fields are referred to as university disciplinary areas. At the bottom level, there are 370 fields, referred to as specific disciplinary sectors, with each specific disciplinary sector being part of a single university disciplinary area. In Italy, each researcher at a university must belong to exactly one specific disciplinary sector. We will return to the Italian classification system of researchers in Sect. 11.5.

Another example of a classification system of researchers is the classification system of the Mendeley reference management tool. This system was used by *Bornmann* and *Haunschild* [11.17] in a proposal for a field-normalized indicator of scientific impact based on Mendeley reader counts. In the Mendeley classification system, a distinction between 28 fields is made. Each Mendeley user is able to assign him- or herself to one of these 28 fields.

11.4 Overview of Field-Normalized Indicators

In this section, we provide an overview of field-normalized indicators that have been proposed in the scientometric literature. The literature on field-normalized indicators is extensive. We, therefore, do not discuss each individual proposal presented in the literature. Instead, our focus is on what we consider to be the more significant contributions that have been made. Other contributions may not be covered or may be mentioned only very briefly. We also do not aim to give a historical overview of the literature. We discuss important ideas presented in the literature but we do not necessarily trace the historical development of these ideas.

In principle, field-normalized variants can be developed for any type of scientometric indicator. In practice, however, scientometricians have put most effort into the development of field-normalized indicators of the impact of scientific publications, where impact is typically operationalized using citations. Our focus in this section is, therefore, mostly on field-normalized indicators of impact, although we also discuss field-normalized indicators of productivity. Most of the indicators that we consider in this section rely on field classification systems such as the ones introduced in the previous section, but we also discuss indicators that do not require a field classification system.

Field-normalized indicators typically normalize not only for the field of a publication but also for the age of a publication. This is important in the case of indicators based on citations, since older publications have had more time to receive citations than younger publications. Indicators may also normalize for other characteristics of a publication. For instance, they sometimes normalize for publication type, where a distinction can be made between categories such as research article, review article, and letter.

11.4.1 Indicators of Impact: Indicators Based on Normalized Citation Scores

The normalized citation score of a publication can be defined in different ways. The most straightforward approach is to define it as the ratio of the actual and the expected number of citations of a publication, where the expected number of citations of a publication equals the average number of citations of all publications in the same field and in the same publication year (and often also in the same publication type category). Whether publications are in the same field is determined based on a field classification system, such as one of the systems discussed in Sect. 11.3.

In order to obtain indicators at the level of, for instance, a research group, a research institution, or a journal, the normalized citation scores of individual publications need to be aggregated. This is typically done either by averaging or by summing the normalized citation scores. Averaging the scores yields a so-called size-independent indicator of impact, while summing the scores gives a size-dependent impact indicator. These indicators are known under various different names. The size-independent indicator is, for instance, known as the mean normalized citation score [11.18], the item-oriented field normalized citation score average [11.19], the category normalized citation impact (in the commercial InCites tool), and the field weighted citation impact (in the commercial Scopus and SciVal tools). The size-dependent indicator is sometimes referred to as the total normalized citation score [11.18].

A recent development is the application of the above approach for calculating field-normalized impact indicators to bookmarks in Mendeley instead of citations. Studies of field-normalized indicators based on Mendeley bookmarks, often interpreted in terms of readership, have been reported by *Fairclough* and *Thelwall* [11.20] and *Haunschild* and *Bornmann* [11.21].

A number of alternative approaches have been explored for defining the normalized citation score of a publication. One alternative is to leave out noncited publications from the calculation of the expected number of citations of a publication [11.22, 23]. Another alternative is to determine the expected number of citations of a publication based on the idea of so-called exchange rates, where the similarity between fields in the shape of citation distributions is used to determine how many citations in one field can be considered equivalent to a given number of citations in another field [11.24, 25]. A third alternative is to apply a logarithmic transformation to the citation counts of publications [11.19, 26–28]. A fourth alternative is to transform citation counts into z-scores [11.29–31]. This approach can be combined with a logarithmic transformation of citation counts [11.19]. A fifth alternative is to transform citation counts using a two-parameter power-law function [11.32]. Finally, a sixth alternative proposed in the literature is to transform citation counts into binary variables based on whether or not publications have been cited [11.28].

There has also been considerable discussion in the literature about the best way to calculate field-normalized impact indicators at aggregate levels, for instance, at the level of research groups or research institutions [11.18, 19, 33–37]. The approach discussed above, in which normalized citation scores of individual publications are averaged or summed, is nowadays the most commonly used approach. An alternative approach is to calculate the average or the sum of the actual citation counts of a set of publications and to divide the outcome by the average of the expected citation counts of the same set of publications [11.38–40]. In this alternative approach, normalization can be considered to take place at the level of an oeuvre of publications rather than at the level of individual publications [11.34]. When an analysis includes publications from multiple fields or multiple years, normalization at the oeuvre level will generally yield results that are different from the outcomes obtained by normalizing at the level of individual publications. We refer to *Larivière* and *Gingras* [11.41], *Waltman* et al. [11.42], and *Herranz* and *Ruiz-Castillo* [11.43] for empirical analyses of the differences between the two approaches.

Another issue in the calculation of field-normalized impact indicators at aggregate levels is the choice of a counting method for handling co-authored publications. Full and fractional counting are the two most commonly used counting methods. In the case of full counting, each publication is fully counted for each co-author. On the other hand, in the case of fractional counting, a publication with n co-authors is counted with a weight of $1/n$ for each co-author. The choice of counting method influences the extent to which an indicator can be considered to provide properly field-normalized statistics [11.44]. We will return to this issue in Sect. 11.5.1.

11.4.2 Indicators of Impact: Indicators Based on Percentiles

Percentile-based impact indicators value publications based on their position in the citation distribution of their field and publication year, where fields are defined using a field classification system, for instance, one of the systems discussed in Sect. 11.3. In the most straightforward case, these indicators make a distinction between lowly and highly cited publications. For instance, all publications that in terms of citations belong to the top 10%, top 5%, or top 1% of their field and publication year may be regarded as highly cited, as suggested by *Tijssen* et al. [11.45] and *van Leeuwen* et al. [11.46]. A generalization of this idea was proposed by *Leydesdorff* et al. [11.47]. In their proposal, a number of classes of publications are distinguished. Each class of publications is defined in terms of percentiles of the citation distribution of a field and publication year. The first class may, for instance, include all publications whose number of citations is below the 50th percentile of the citation distribution of their field and publication year, the second class may include all publications whose number of citations is between the 50th and the 75th percentile, and so on. In the proposed approach, publications are valued based on the class to which they belong. Publications in the lowest class have a value of 1, publications in the second-lowest class have a value of 2, and so on.

A difficulty in the calculation of percentile-based indicators is the issue of ties, that is, multiple publications with the same number of citations. Suppose we want to identify the 10% most frequently cited publications in a certain field and publication year. We then need to find a threshold such that exactly 10% of the publications in this field and publication year have a number of citations that is above the threshold. In practice, it will usually not be possible to find such a threshold. Because of the issue of ties, typically, any threshold will yield either too many or too few publications whose number of citations is above the threshold. This means that fields cannot be made fully comparable, since the distortion caused by the issue of ties will be different in different fields. In the literature, various approaches for dealing with the issue of ties have been explored [11.46–49]. We refer to *Waltman* and *Schreiber* [11.49] for a summary of these approaches and to *Schreiber* [11.50] for an empirical comparison.

Field-normalized impact indicators can also be constructed by combining the idea of percentile-based indicators with the idea of indicators based on normalized citation scores. Such an approach was introduced by *Albarrán* et al. [11.51, 52]. In the proposed approach, indicators are used to characterize the distribution of citations over the highly cited publications in a field. The indicators resemble indicators developed in the field of economics for characterizing income distributions.

Glänzel [11.53] and *Glänzel* et al. [11.54] proposed indicators that, like the above-mentioned indicators proposed by *Leydesdorff* et al. [11.47], distinguish between a number of classes of publications. However, instead of percentiles, these indicators rely on the method of characteristic scores and scales [11.55] to define the classes. Publications belong to the lowest class if they have fewer citations than the average of their field, they belong to the second-lowest class if they do not belong to the lowest class and if they have fewer citations than the average of all publications not belonging to the lowest class, and so on. An alternative approach is to define the classes based on median instead of average citation counts [11.56].

Percentile-based approaches may also be used to normalize altmetric indicators. *Bornmann* and *Haunschild* [11.57] suggested a percentile-based approach for normalizing Twitter counts.

11.4.3 Indicators of Impact: Indicators that Do Not Use a Field Classification System

All field-normalized indicators discussed so far rely on a field classification system that operationalizes scientific fields. As discussed in Sects. 11.2 and 11.3, the operationalization of fields is a difficult problem. Field classification systems offer a simplified representation of fields. By necessity, any field classification system relies partly on arbitrary and contestable choices. In this section, we discuss field-normalized impact indicators with the attractive property that they do not require a field classification system.

An approach that has been explored in the literature is to identify for each publication a set of similar publications, allowing the citation score of the focal publication to be compared with the citation scores of the identified similar publications. Similar publications may be identified based on shared references (i. e., bibliographic coupling relations), as suggested by *Schubert* and *Braun* [11.58, 59]. Alternatively, as demonstrated by *Colliander* [11.60], the identification of similar publications may be done based on a combination of shared references and shared terms. Another possibility is to use co-citation relations to identify similar publications. This idea is used in the relative citation ratio indicator, an indicator introduced by a research team at the US National Institutes of Health [11.61] that has attracted a significant amount of attention. We refer to *Janssens* et al. [11.62] for a critical discussion of the relative citation ratio indicator (for a response by the original authors, see [11.63]). Instead of working at the level of individual publications, it is also possible to work at the journal level. The citation score of a journal can then be compared with the citation scores of other similar journals. The latter journals may be identified based on citations given to the focal journal [11.64].

Another field normalization approach that does not require a field classification system is known as citing-side normalization [11.65], sometimes also referred to as fractional citation weighting [11.65], fractional citation counting [11.66], source normalization [11.67], or a priori normalization [11.68]. Citing-side normalization is based on the idea that differences between fields in citation density are to a large extent caused by the fact that in some fields publications tend to have longer reference lists than in other fields. Citing-side normalization performs a correction for the length of the reference list of citing publications. The basic idea of citing-side normalization can be implemented in different ways. One possibility is to correct for the average reference list length of citing journals [11.65, 69]. Another possibility is to correct for the reference list length of individual citing publications [11.66–68, 70, 71]. A combination of these two options is possible as well, and this is how citing-side normalization is implemented in the current version of the Source Normalized Impact per Paper (SNIP) journal impact indicator [11.72].

Instead of correcting for reference list length on the citing side, an alternative approach is to correct for reference list length on the cited side. In this approach, a correction can be made for either the reference list length of a cited publication [11.73] or the average reference list length of a cited journal [11.74–76]. A third possibility is to correct for the average reference list length of all publications belonging to the same field as a cited publication [11.77]. However, this again requires a field classification system, just like in the case of the indicators discussed in Sects. 11.4.1 and 11.4.2.

Recursive impact indicators, first introduced by *Pinski* and *Narin* [11.78] and often inspired by the well-known PageRank algorithm [11.79], offer another approach that is related to the idea of citing-side normalization. Examples of recursive impact indicators are the eigenfactor and article influence indicators of journal impact [11.80, 81] and the SCImago journal rank (SJR) indicator [11.82, 83]. We refer to *Waltman* and *Yan* [11.84] and *Fragkiadaki* and *Evangelidis* [11.85] for overviews of the literature on recursive impact indicators and to *Waltman* and *van Eck* [11.86] for a discussion of the relation between these indicators and citing-side normalized indicators.

11.4.4 Indicators of Productivity

Although field-normalized indicators of impact have received most attention in the scientometric literature, some attention has also been given to field-normalized indicators of productivity (sometimes also referred to as efficiency). Productivity indicators can, for instance, be calculated for researchers, research groups, and research institutions. A simple productivity indicator is the average number of publications produced per researcher. A more advanced productivity indicator may also take into account the number of citations publications have received. Field-normalized productivity indicators perform a correction for differences between fields in the rate at which publications are produced and citations are received.

Field-normalized productivity indicators play a prominent role in the work of *Giovanni Abramo* and *Ciriaco Andrea D'Angelo*. In particular, *Abramo* and *D'Angelo* make extensive use of an indicator referred to as the fractional scientific strength (FSS). For an individual researcher, FSS essentially equals the sum of the normalized citation scores (Sect. 11.4.1) of the publications of the researcher divided by the salary of the researcher. Likewise, for a group of researchers working in the same field, FSS equals the sum of the normalized citation scores of their publications divided by their total salary. When FSS is calculated for a group of researchers working in different fields, for instance, all researchers affiliated with a particular research institution, a correction needs to be made for differences between fields in the average publication output and the average salary of researchers. One way in which this can be done is by first calculating each researcher's field-normalized FSS, defined as the researcher's FSS divided by the average FSS of all researchers working in the same field and then calculating the average field-normalized FSS of all researchers. We refer to *Abramo* and *D'Angelo* [11.16] for a more detailed discussion of the calculation of the FSS indicator. For a discussion of an alternative productivity indicator, based on highly cited publications instead of normalized citation scores, we refer to *Abramo* and *D'Angelo* [11.87].

In practice, calculating the FSS indicator is highly challenging because it requires data on the publications and the salaries of all researchers working in a field. *Abramo* and *D'Angelo* address this difficulty by taking into account only Italian publications and Italian researchers in the calculation of the FSS indicator. In Italy, unlike in most other countries, the data required for the calculation of the FSS indicator is available. *Abramo* and *D'Angelo* calculate normalized citation scores of publications using the WoS journal-based field classification system (Sect. 11.3.1). However, they also need a second classification system. To calculate researchers' field-normalized FSS, they rely on a classification system of Italian researchers (Sect. 11.3.3).

In most countries, the data needed to calculate field-normalized productivity indicators is not available. Obtaining productivity indicators that allow for meaningful cross-country comparisons is even more challenging, as pointed out by *Aksnes* et al. [11.88]. An interesting proposal for calculating field-normalized productivity indicators, even when only limited data is available, was presented by *Koski* et al. [11.89]. This proposal focuses on the difficulty of researchers that have no publications in a certain time period. These researchers are invisible in databases such as WoS and Scopus, which causes problems when using these databases to calculate productivity indicators. To deal with this issue, a statistical methodology is proposed for estimating the number of researchers without publications.

A number of studies have focused specifically on designing field-normalized indicators of the productivity of individual researchers. In particular, several proposals have been made for variants of the h-index

[11.90] that correct for field differences [11.91–95]. Other interesting proposals for comparing individual researchers active in different fields were presented by *Kaur* et al. [11.96] and *Ruocco* and *Daraio* [11.97].

It is important to be aware of the difference between productivity indicators and size-independent impact indicators. Both types of indicators are independent of size, which is convenient, for instance, when making comparisons between larger and smaller research institutions. However, the two types of indicators are based on fundamentally different notions of size. Productivity indicators take an input perspective on the notion of size, for instance, the number of researchers affiliated with an institution. Size-independent impact indicators take an output perspective on the notion of size, namely the number of publications produced by an institution. From a conceptual point of view, for many purposes the input perspective seems preferable over the output perspective. From a practical point of view, however, taking the input perspective often is not possible because the data required is not available. A more elaborate discussion of the pros and cons of productivity indicators and size-independent impact indicators can be found in a recent special section of *Journal of Informetrics* [11.98]. In this special section, a discussion paper by *Abramo* and *D'Angelo* [11.99] argues in favor of the use of productivity indicators, while other contributions defend the use of size-independent impact indicators. We refer to *Abramo* and *D'Angelo* [11.100] for an institutional-level comparison between productivity indicators and size-independent impact indicators.

11.5 Evaluation of Field-Normalized Indicators

The discussion in the previous section has shown that a large variety of field-normalized indicators have been proposed in the literature. This, of course, raises various questions: Do the indicators discussed in the previous section indeed provide properly field-normalized statistics? What are the advantages and disadvantages of the different ways in which field normalization can be performed? Is it possible to identify one specific approach to field normalization that can be considered superior over other approaches? To provide some partial answers to these questions, we now discuss the scientometric literature on the evaluation of field-normalized indicators. We restrict the discussion to indicators of impact.

To evaluate field-normalized indicators, some scientometricians choose to analyze the theoretical properties of indicators, while other scientometricians prefer to study the empirical characteristics of indicators. Different approaches to evaluate field-normalized indicators sometimes lead to different conclusions. For instance, from a theoretical perspective, an indicator may seem appealing, while from an empirical perspective the same indicator may not seem very attractive. Below, we first discuss the theoretical evaluation of field-normalized indicators. We then turn to empirical evaluation.

11.5.1 Theoretical Evaluation of Indicators

In theoretical approaches to the evaluation of field-normalized indicators, the theoretical properties of indicators are studied. These are properties that do not depend on empirical data based on which indicators are calculated. After the theoretical properties of indicators have been established, the indicators are evaluated by deciding whether or not their properties are considered desirable. Whether a certain property is desirable is a subjective question that may legitimately be answered differently by different people. Theoretical evaluation, therefore, does not offer a universal and definitive answer to the question of whether one indicator is superior over another. Instead, it aims to provide a deep understanding of the key differences between indicators. This may then guide users in choosing the indicator that best serves their needs.

In the calculation of the normalized citation score of a publication, defined as the ratio of the actual and the expected number of citations of a publication (Sect. 11.4.1), theoretical considerations may help to choose between different ways in which the expected number of citations of a publication can be defined. The most common approach is to define a publication's expected citation count as the average citation count of all publications in the same field and in the same publication year. An alternative approach is to consider in this definition only publications that have been cited at least once [11.22, 23]. In the case of the former approach, for each combination of a field and a publication year, the average normalized citation score of all publications in that field and publication year equals exactly 1. This may be regarded as an important property for a field-normalized indicator. The approach in which non-cited publications are left out from the definition of a publication's expected citation count does not have this property, which may be seen as a disadvantage of this approach.

Another issue in the calculation of the normalized citation score of a publication is the way in which publications belonging to multiple fields are handled. Based on the idea that the average normalized citation score of all publications in a field and publication year should equal 1, it can be argued that the expected citation count of a publication belonging to multiple fields should be defined as the harmonic average of the expected citation counts corresponding to the different fields [11.18]. However, a theoretical analysis presented by *Smolinsky* [11.101] showed that there are also other ways in which the expected citation count of a publication belonging to multiple fields can be defined. These alternative approaches lead to additional properties that may be considered attractive, but they have the disadvantage of introducing challenging computational issues.

As discussed in Sect. 11.4.1, there are different ways in which field-normalized indicators can be calculated at the aggregate level of, for instance, a research institution. The oeuvre argument of *Moed* [11.34] is a theoretical argument in favor of one approach, while the consistency argument of *Waltman* et al. [11.18] is a theoretical argument in favor of another approach. According to the oeuvre argument, it should not make a difference whether a citation is given to one publication in the oeuvre of a research unit or to some other publication in the same oeuvre. The basic idea of the consistency argument is that the ranking of two research units relative to each other should not change when both units make the same performance improvement. The oeuvre and consistency arguments can also be used to characterize some of the key differences between two versions of the SNIP journal impact indicator [11.67, 72].

The choice of the counting method used to handle co-authored publications in the calculation of a field-normalized indicator can also be analyzed theoretically. When the full counting method is used, each publication is fully counted for each co-author, as explained in Sect. 11.4.1. On the other hand, when using a fractional counting method, co-authored publications are counted with a lower weight than publications that have not been co-authored. As pointed out by *Waltman* and *van Eck* [11.44], in the case of fractional counting, the mean normalized citation score indicator (Sect. 11.4.1) has the property that the average value of the indicator for all research institutions active in a field equals exactly 1. In the case of full counting, the indicator does not have this property. Using the full counting method, co-authored publications are counted multiple times, once for each of the co-authors. This double counting of co-authored publications, which tend to be publications that have received relatively large numbers of citations, has an inflationary effect. It typically causes the mean normalized citation score indicator to have an average value for all research institutions active in a field that is above 1. Because of this inflanatory effect, which is larger in some fields than in others, the full counting method provides statistics that are only partly field normalized. In order to obtain properly field-normalized statistics, a fractional counting method needs to be used. Alternatively, the use of a so-called multiplicative counting method [11.102] can be considered.

11.5.2 Empirical Evaluation of Indicators

Empirical approaches to the evaluation of field-normalized impact indicators focus on three questions. First, assuming that a certain field classification system offers a satisfactory representation of scientific fields, which field-normalized indicators provide the best normalization? Second, to what extent do different field classification systems offer good representations of scientific fields, in particular for the purpose of field normalization? Third, which field-normalized indicators have the strongest correlation with peer review?

The idea of universality of citation distributions plays a key role in the literature dealing with the first question. Citation distributions are considered to be universal if the distribution of normalized citation scores is essentially identical for all scientific fields. The idea of universality of citation distributions was introduced by *Radicchi* et al. [11.95]; see also [11.10]. They claimed that universality of citation distributions can be achieved using a straightforward normalization approach in which the number of citations of each publication in a field is divided by the average number of citations of all publications in the field (excluding noncited publications). However, in subsequent studies, it has been shown that this straightforward normalization approach yields citation distributions that are only approximately universal [11.29, 103, 104].

Based on the idea of universality of citation distributions, a so-called fairness test for field-normalized indicators was proposed [11.105]. This test has been used in various studies in which field-normalized indicators are compared [11.32, 105, 106]. The objective of having normalized citation distributions that are identical across fields also serves as the foundation of a methodology for quantifying the degree to which field-normalized indicators succeed in correcting for field differences [11.25]. This methodology has been applied in various studies [11.24, 25, 107, 108]. Using the methodology of *Crespo* et al. [11.25], it was found that the normalization approach proposed by *Radicchi* and *Castellano* [11.32], based on a two-parameter power law transformation of citation counts, outperforms a number of other normalization approaches [11.107]. However, the standard approach of dividing the actual number of citations of a publication by the expected number of citations has also been shown to perform well. A study by *Abramo* et al. [11.22] in which a comparison is made of a number of field-normalized indicators is also based on the idea of trying to obtain normalized citation distributions that are identical across fields.

The above-mentioned studies assume that one has a satisfactory field classification system. They do not evaluate whether a certain classification system offers a good representation of scientific fields. This limited perspective on the evaluation of field-normalized indicators was criticized by *Sirtes* [11.109] in a letter commenting on *Radicchi* and *Castellano* [11.105]; for a response, see [11.110]. According to *Sirtes* [11.109]; see also [11.108], it is incorrect to evaluate a field-normalized indicator using the same classification system that is also used in the calculation of the indicator. This brings us to the second question raised in the beginning of this section: How suitable are different field classification systems for the purpose of field normalization?

Evaluations of the use of the WoS journal-based field classification system for the purpose of field normalization have been reported by *van Eck* et al. [11.111] and *Leydesdorff* and *Bornmann* [11.112]. In both studies, the appropriateness of the fields in the WoS classification system for normalization purposes is questioned. For other studies questioning the use of the WoS classification system and proposing the use of alternative classification systems, we refer to *Bornmann* et al. [11.8], *Neuhaus* and *Daniel* [11.9], *van Leeuwen* and *Calero Medina* [11.11], and *Ruiz-Castillo* and *Waltman* [11.13]. A systematic methodology for comparing the suitability of different classification systems for field-normalization purposes was presented by *Li* and *Ruiz-Castillo* [11.113]. We refer to *Perianes-Rodriguez* and *Ruiz-Castillo* [11.14] for an application of this methodology.

Empirical approaches to the evaluation of field-normalized impact indicators also study the extent to which these indicators correlate with peer review. At the level of research programs and research departments in the natural sciences, indicators that use the standard normalization approach of dividing the actual number of citations of a publication by the expected number of citations have been shown to be moderately correlated with peer review assessments made by expert committees [11.114, Chap. 19], [11.115]. The correlation between normalized impact indicators and peer review has also been analyzed based on peer review outcomes from the Research Assessment Exercise (RAE) in the UK [11.116]. The main finding of this analysis is that impact indicators normalized at the level of journals hardly correlate with peer review, while impact indicators normalized at the level of journal-based fields in the WoS database or units of assessment in the RAE correlate significantly with peer review.

At the level of individual publications, the recently introduced relative citation ratio indicator has been claimed to be *well correlated* with expert judgments [11.61, p. 9]. However, in a study by *Bornmann* and *Haunschild* [11.117], the correlation between the relative citation ratio indicator and expert judgments was characterized as *only low to medium* (p. 1064). In addition, it has been shown that, in terms of correlation with peer review, the relative citation ratio indicator has a performance that is similar to other field-normalized impact indicators. The studies by *Hutchins* et al. [11.61] and *Bornmann* and *Haunschild* [11.117] both make use of F1000 post-publication peer review data. Data from F1000 has also been used to analyze, at the level of individual publications, how strongly a number of field-normalized impact indicators correlate with peer review [11.118]. It was found that different field-normalized impact indicators all have a similar correlation with peer review. However, the authors leave open the possibility that F1000 data may not be sufficiently accurate to make fine-grained distinctions between different field-normalized impact indicators.

11.6 How Much Difference Does It Make in Practice?

We have discussed a large number of field-normalized indicators as well as a large number of field classification systems that can be used by these indicators. We now consider the following question: How much difference does the choice of a field-normalized indicator, and possibly also a field classification system, make in practice, for instance, when field-normalized indicators are used in the evaluation of research institutions, research groups, or individual researchers?

Various papers have presented analyses that provide insight into this question, most of them focusing on field-normalized indicators of impact. Before reporting our own analysis, we first briefly mention some of these papers, without going into the details of their findings. At the level of individual publications, the sensitivity of field-normalized indicators to the choice of a field classification system was studied by *Zitt* et al. [11.3]. In the context of quantifying the impact of journals, field-normalized indicators that use a field classification system (Sects. 11.4.1 and 11.4.2) were compared with field-normalized indicators that use citing-side normalization (Sect. 11.4.3) and that do not require a field classification system [11.106, 119]. Similar comparisons have also been made for indicators based on Mendeley bookmarks [11.17]. In the context of quantifying the impact of research institutions and their internal units, a number of studies investigated for specific field-normalized indicators the effect of the choice of a field classification system. The use of the WoS journal-based classification system was compared with the use of other less fine-grained journal-based systems [11.38, 120], but also with the use of more fine-grained publication-based systems [11.13]. In addition to analyzing the effect of the choice of a classification system, studies have also compared different normalization approaches for a given classification system. *Perianes-Rodriguez* and *Ruiz-Castillo* [11.121], for instance, performed a comparison of two different ways in which normalized impact indicators can be obtained at the aggregate-level of research institutions. Finally, as has already been mentioned, field-normalized indicators of productivity have received relatively limited attention in the literature. A comparison of two ways in which the FSS indicator (Sect. 11.4.4) can be calculated at the level of research institutions was reported by *Abramo* and *D'Angelo* [11.122].

11.6.1 Empirical Analysis of the Sensitivity of Field-Normalized Impact Indicators to the Choice of a Field Classification System

Complementary to the studies mentioned above, we now present our own analysis. Our focus is on the sensitivity of field-normalized impact indicators to the choice of a field classification system. We are interested in particular in the sensitivity of the indicators at lower levels of aggregation, that is, at the level of internal units within a research institution. This level is highly relevant in practical applications of field-normalized impact indicators.

The mean normalized citation score (MNCS) [11.18] (Sect. 11.4.1), and the proportion of top 10% publications (PP(top 10%)) [11.49] (Sect. 11.4.2), represent two of the most frequently used size-independent field-normalized impact indicators (taking into account also the use of variants of these indicators in commercial tools such as InCites and SciVal). Given the popularity of these indicators, it is important to understand their sensitivity to the choice of a field classification system. In this section, we, therefore, analyze the sensitivity of these indicators to the choice between, on the one hand, a traditional journal-based classification system, namely the classification system consisting of about 250 fields that is available in the WoS database, and, on the other hand, a publication-based classification system constructed algorithmically using the methodology of *Waltman* and *van Eck* [11.12]. At the level of research institutions, the sensitivity to the choice between these two classification systems has been found to be relatively limited [11.13]. However, this sensitivity has not yet been analyzed in a systematic way for smaller units. Below we present such an analysis for internal units within a large European university.

Our analysis is based on the WoS database. More specifically, we use the Science Citation Index Expanded, the Social Sciences Citation Index, and the Arts & Humanities Citation Index. We use data on the publications of our focal university in the period 2010–2014. Publications are assigned to internal units within the university at three hierarchical levels. We refer to these levels as the faculty level, the department level, and the research group level. We take into only take into account units that have at least 50 publications. Also, only publications classified as research articles or review articles are considered. There are 13 faculties, 36 departments, and 130 research groups with 50 or more publications. Some basic statistics on the numbers of publications of these faculties, departments, and research groups are reported in Table 11.1.

Citations are counted until the end of 2015. Author self-citations are excluded. In the case of the WoS journal-based classification system, publications in journals belonging to the *Multidisciplinary Sciences* category are reassigned to other categories based on their references. In the case of the publication-based classification system, we use a system that includes about 4000 fields. This is in line with the recommendation made by *Ruiz-Castillo* and *Waltman* [11.13]. The calculation of the

Table 11.1 Statistics on the numbers of publications of the faculties, departments, and research groups of the focal university

	Number of units	Number of publications			
		Min.	Max.	Mean	Median
Faculties	13	88	7626	1423.7	945.0
Departments	36	54	2785	560.3	322.5
Research groups	130	50	766	166.4	116.5

MNCS and PP(top 10%) indicators is based on, respectively, *Waltman* et al. [11.18] and *Waltman* and *Schreiber* [11.49]. Normalization is performed for field and publication year, but not for publication type. A full counting approach is taken. Hence, each publication authored by a unit is fully counted for that unit, irrespective of possible co-authorship with other units inside or outside the focal university.

The results of the analysis are presented in Fig. 11.1 for the 13 faculties, in Fig. 11.2 for the 36 departments, and in Fig. 11.3 for the 130 research groups. Each figure shows two scatter plots, one for the MNCS indicator

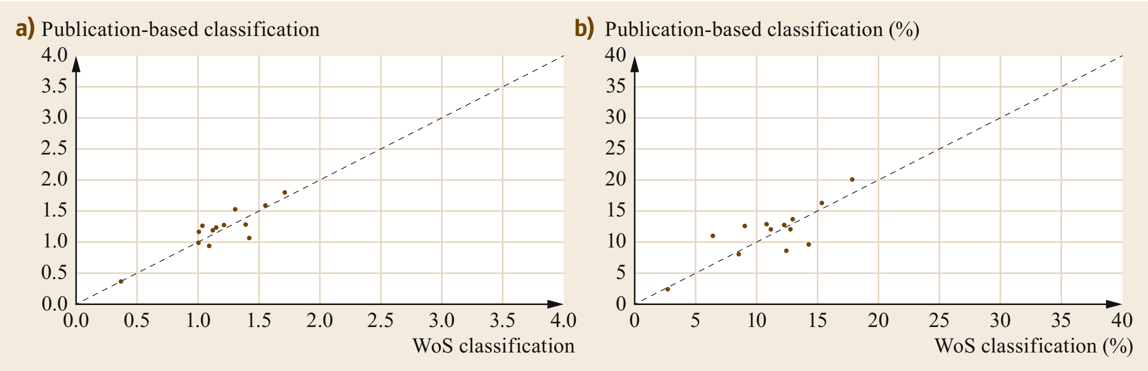


Fig. 11.1a,b Scatter plots of the MNCS **(a)** and PP(top 10%) **(b)** values of the 13 faculties of the focal university, obtained using either the WoS journal-based classification system or a publication-based classification system

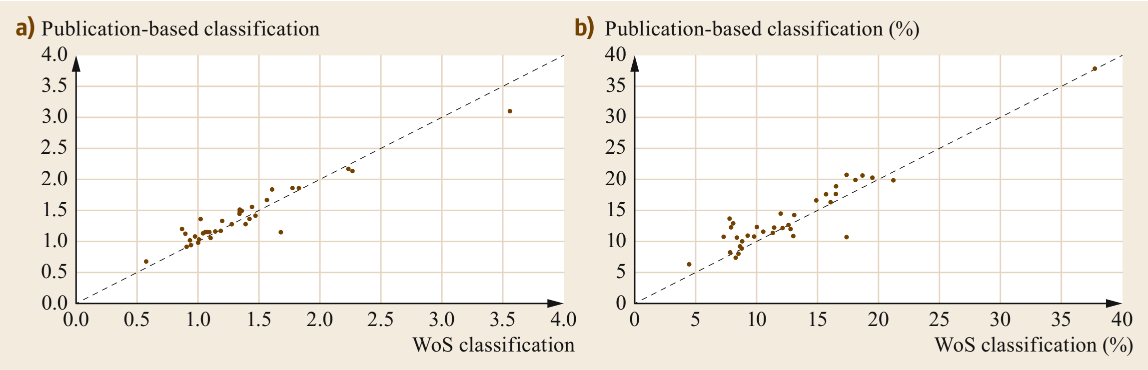


Fig. 11.2a,b Scatter plots of the MNCS **(a)** and PP(top 10%) **(b)** values of the 36 departments of the focal university, obtained using either the WoS journal-based classification system or a publication-based classification system

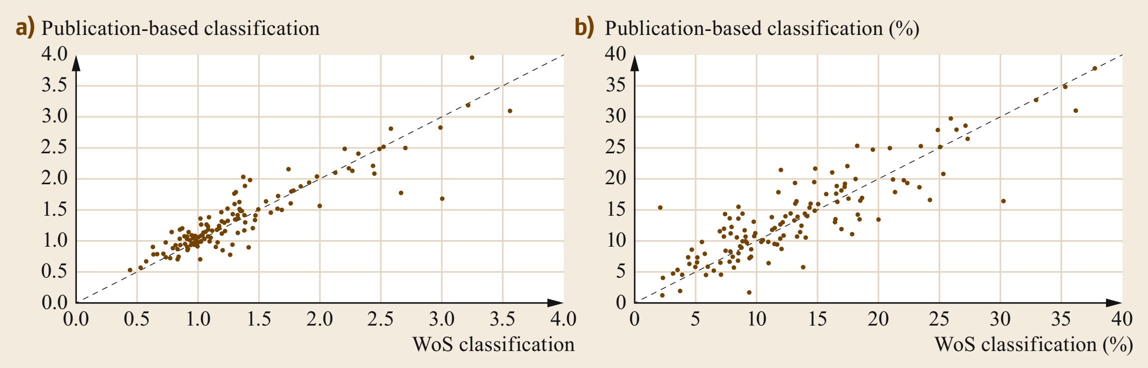


Fig. 11.3a,b Scatter plots of the MNCS **(a)** and PP(top 10%) **(b)** values of the 130 research groups of the focal university, obtained using either the WoS journal-based classification system or a publication-based classification system

Table 11.2 Statistics on the differences between the indicator values obtained using the WoS journal-based classification system and the indicator values obtained using a publication-based classification system

	Pearson correlation		Mean absolute difference		% units with large difference	
	MNCS	PP(top 10%)	MNCS	PP(top 10%)	MNCS	PP(top 10%)
Faculties	0.89	0.80	0.12	2.0%	0.0%	0.0%
Departments	0.95	0.93	0.12	1.8%	2.8%	5.6%
Research groups	0.91	0.87	0.17	2.7%	5.4%	14.6%

and one for the PP(top 10%) indicator. In addition, Table 11.2 reports a number of statistics that summarize the differences between the results obtained using the WoS journal-based classification system and those obtained using the publication-based classification system. For both the MNCS indicator and the PP(top 10%) indicator, the table presents the Pearson correlation between the results obtained using the two classification systems. Moreover, the table also shows the mean absolute difference between the results and the percentage of units for which the difference is considered to be large. A difference in the MNCS value of a unit of more than 0.5 is regarded as large. In the case of the PP(top 10%) indicator, we regard a difference of more than 5 percentage points as large.

The results in Table 11.2 show that the mean absolute differences are larger at the level of the research groups than at the level of the faculties and the departments. Likewise, the percentage of units with large differences is highest at the research group level. These findings may not be surprising. Research groups on average have a much smaller number of publications than faculties and departments (Table 11.1), and, therefore, the MNCS and PP(top 10%) values of research groups can be expected to be more sensitive to the choice of a classification system than the corresponding values of faculties and departments. Based on the results in Table 11.2 and Figs. 11.1–11.3, it can also be concluded that the PP(top 10%) indicator is more sensitive to the choice of classification than the MNCS indicator.

Based on our results, how sensitive are field-normalized impact indicators to the choice of a field classification system? The answer to this question may depend on the expectations that one has. Some readers may consider the differences between the results obtained using the WoS journal-based classification system and the results obtained using the publication-based classification system to be within an acceptable margin. Others may be concerned to see, for instance, that for about one out of seven research groups the PP(top 10%) indicator increases or decreases by more than 5 percentage points when changing the classification system based on which the indicator is calculated (Table 11.2). Our perspective is that the results illustrate the risk of overinterpreting field-normalized indicators, especially at lower levels of aggregation, such as the research group level. There is no perfect way to correct for differences between fields. Different field normalization approaches make different choices in how they correct for field differences. Each approach is informative in its own way. When working with one specific field-normalization approach, it is essential to keep in mind that this approach offers just one perspective on field normalization and that other approaches will give a different perspective, in some cases even a perspective that may be different in a quite fundamental way.

11.7 Conclusion

Some critics question whether field normalization is truly attainable. In the literature, this viewpoint is represented by *Kostoff* [11.123] and *Kostoff* and *Martinez* [11.124], who criticize the idea of field-normalized impact indicators, arguing that citation counts of publications should be compared only if publications are very similar to each other. According to *Kostoff* and *Martinez* [11.124, p. 61]:

> a meaningful 'discipline' citation average may not exist, and the mainstream large-scale mass production semi-automated citation analysis comparisons may provide questionable results.

In principle, critics make a valid point. Taking their position to the extreme, one could argue that every publication is unique in its own way and, consequently, that any comparison of citation counts of publications is problematic. Likewise, it could be argued that every researcher is unique and that any comparison of publication and citation counts of researchers is, therefore, in some sense unfair.

However, one may also take a more pragmatic perspective on the idea of field normalization. In managing and evaluating scientific research, there is often a need to compare different research units (e. g., research institutions, research groups, or individual researchers).

Scientometric indicators, of course, provide an incomplete picture of the units to be compared. Moreover, these indicators are affected by all kinds of distorting factors, for instance, related to the characteristics of the underlying data sources, the peculiarities of the units to be compared, and the nature of the scientific fields in which these units are active. Nevertheless, despite their limitations, scientometric indicators provide useful and relevant information for supporting the management and evaluation of scientific research. In many cases, the usefulness of scientometric indicators can be increased by making corrections for some of the most significant distorting factors, and field differences typically are one such a factor. Field normalization does not correct for all distorting factors, but it corrects at least partly for one of the most important ones. From this point of view, field normalization serves an important practical purpose.

11.7.1 Strengths and Weaknesses of Different Field-Normalization Approaches

We have provided an overview of a large number of approaches to field normalization. Although some field-normalization approaches can be considered superior over others, we do not believe there to be a single optimal approach. Instead, there is a trade-off between the strengths and weaknesses of different approaches. Some field-normalization approaches have a high level of technical sophistication. These approaches may, for instance, use an algorithmically constructed publication-based field classification system or they may not need a classification system at all, and instead of the traditional full counting method, these approaches may use a fractional counting method for dealing with co-authored publications. Other field normalization approaches are much more basic. For instance, they rely on the standard journal-based classification system made available in a database such as WoS or Scopus and they handle co-authored publications using the standard full counting method. In general, the more sophisticated approaches can be expected to better correct for field differences than the more basic approaches. On the other hand, however, the more basic approaches tend to be easier to understand and more transparent. This enables users to carefully reflect on what a field-normalized indicator does and does not tell them, and it allows users to recognize the limitations of the indicator. The more sophisticated approaches tend to be black boxes for many users, forcing users to blindly trust the outcomes provided by these approaches. Due to the low level of transparency, it is difficult for users to understand the limitations of the more sophisticated approaches and to interpret the outcomes obtained using these approaches in the light of these limitations.

As a general rule, in situations in which in-depth reflection on scientometric indicators is desirable or even essential, for instance, when indicators are used to support the evaluation of individual researchers, we recommend the use of simple and transparent field-normalization approaches. Complex non-transparent approaches should not be used in such situations. On the other hand, there are also situations in which the use of more advanced field-normalization approaches, possibly with a relatively low level of transparency, may be preferable. This could be the case in situations in which scientometric indicators are used at a high level of aggregation, for instance, at the level of entire research institutions or countries, where in-depth reflection on the indicators may hardly be possible, or in situations in which scientometric indicators are used in a purely algorithmic way, for instance, when they are embedded in a funding allocation model.

11.7.2 Contextualization as an Alternative Way to Deal with Field Differences

We end this chapter by pointing out that field normalization is not the only way to deal with field differences in scientometric analyses. When detailed assessments need to be made at the level of individual researchers or research groups, an alternative approach is to use straightforward non-normalized indicators and to contextualize these indicators with additional information that enables evaluators to take into account the effect of field differences [11.125]. For instance, to compare the productivity of researchers working in different fields, one could present non-normalized productivity indicators (e. g., total publication or citation counts during a certain time period) for each of the researchers to be compared. One could then contextualize these indicators by selecting for each researcher a number of relevant peers working in the same field and by also presenting the productivity indicators for these peers. In this way, each researcher's productivity can be assessed in the context of the productivity of a number of colleagues who have a reasonably similar scientific profile.

An advantage of the above contextualization approach could be that it may lead to a less mechanistic way of dealing with field differences. In our experience, field-normalized indicators tend to be used quite mechanistically, with little attention being paid to their limitations. This is problematic, especially at lower levels of aggregation, for instance, at the level of individual researchers or research groups, where field-normalized indicators are quite sensitive to methodological choices, such as the choice of a field classification system. In a research evaluation, the contextualization approach outlined above may encourage evaluators to reflect more deeply on the effect of field differences and to perform inter-field comparisons in a more cautious and thoughtful way. It may also invite evaluators to combine scientometric evidence of field differences with their own expert knowledge of publication, collaboration, and citation practices in different fields of science. Hence, peer review and scientometrics may be used together in a more integrated manner, which can be expected to improve the way in which research is evaluated [11.125].

Acknowledgments. The authors would like to thank Lutz Bornmann, Robin Haunschild, Loet Leydesdorff, Javier Ruiz-Castillo, and an anonymous referee for their helpful comments on an earlier version of this chapter. These comments have led to numerous improvements.

References

11.1 L. Waltman: A review of the literature on citation impact indicators, J. Informetr. **10**(2), 365–391 (2016)
11.2 C.R. Sugimoto, S. Weingart: The kaleidoscope of disciplinarity, J. Doc. **71**(4), 775–794 (2015)
11.3 M. Zitt, S. Ramanana-Rahary, E. Bassecoulard: Relativity of citation performance and excellence measures: From cross-field to cross-scale effects of field-normalisation, Scientometrics **63**(2), 373–401 (2005)
11.4 Q. Wang, L. Waltman: Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus, J. Informetr. **10**(2), 347–364 (2016)
11.5 W. Glänzel, A. Schubert: A new classification scheme of science fields and subfields designed for scientometric evaluation purposes, Scientometrics **56**(3), 357–367 (2003)
11.6 É. Archambault, O.H. Beauchesne, J. Caruso: Towards a multilingual, comprehensive and open scientific journal ontology. In: *Proc. 13th Int. Conf. Int. Soc. Sci. Informetr., Durban, South Africa*, ed. by E.C.M. Noyons, P. Ngulube, J. Leta (2011) pp. 66–77
11.7 W. Glänzel, A. Schubert, H.J. Czerwon: An item-by-item subject classification of papers published in multidisciplinary and general journals using reference analysis, Scientometrics **44**(3), 427–439 (1999)
11.8 L. Bornmann, R. Mutz, C. Neuhaus, H.D. Daniel: Citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results, Ethics Sci. Env. Polit. **8**(1), 93–102 (2008)
11.9 C. Neuhaus, H.D. Daniel: A new reference standard for citation analysis in chemistry and related fields based on the sections of Chemical Abstracts, Scientometrics **78**(2), 219–229 (2009)
11.10 F. Radicchi, C. Castellano: Rescaling citations of publications in physics, Phys. Rev. E **83**(4), 046116 (2011)
11.11 T.N. van Leeuwen, C. Calero Medina: Redefining the field of economics: Improving field normalization for the application of bibliometric techniques in the field of economics, Res. Eval. **21**(1), 61–70 (2012)
11.12 L. Waltman, N.J. van Eck: A new methodology for constructing a publication-level classification system of science, J. Am. Soc. Inf. Sci. Technol. **63**(12), 2378–2392 (2012)
11.13 J. Ruiz-Castillo, L. Waltman: Field-normalized citation impact indicators using algorithmically constructed classification systems of science, J. Informetr. **9**(1), 102–117 (2015)
11.14 A. Perianes-Rodriguez, J. Ruiz-Castillo: A comparison of the Web of Science and publication-level classification systems of science, J. Informetr. **11**(1), 32–45 (2017)
11.15 S.E. Hug, M. Ochsner, M.P. Brändle: Citation analysis with Microsoft Academic, Scientometrics **111**(1), 371–378 (2017)
11.16 G. Abramo, C.A. D'Angelo: How do you define and measure research productivity?, Scientometrics **101**(2), 1129–1144 (2014)
11.17 L. Bornmann, R. Haunschild: Normalization of Mendeley reader impact on the reader-and paper-side: A comparison of the mean discipline normalized reader score (MDNRS) with the mean normalized reader score (MNRS) and bare reader counts, J. Informetr. **10**(3), 776–788 (2016)
11.18 L. Waltman, N.J. van Eck, T.N. van Leeuwen, M.S. Visser, A.F.J. van Raan: Towards a new crown indicator: Some theoretical considerations, J. Informetr. **5**(1), 37–47 (2011)
11.19 J. Lundberg: Lifting the crown – citation *z*-score, J. Informetr. **1**(2), 145–154 (2007)
11.20 R. Fairclough, M. Thelwall: National research impact indicators from Mendeley readers, J. Informetr. **9**(4), 845–859 (2015)
11.21 R. Haunschild, L. Bornmann: Normalization of Mendeley reader counts for impact assessment, J. Informetr. **10**(1), 62–73 (2016)
11.22 G. Abramo, T. Cicero, C.A. D'Angelo: Revisiting the scaling of citations for research assessment, J. Informetr. **6**(4), 470–479 (2012)
11.23 G. Abramo, T. Cicero, C.A. D'Angelo: How important is choice of the scaling factor in standardizing citations?, J. Informetr. **6**(4), 645–654 (2012)

11.24 J.A. Crespo, N. Herranz, Y. Li, J. Ruiz-Castillo: The effect on citation inequality of differences in citation practices at the Web of Science subject category level, J. Assoc. Inf. Sci. Technol. **65**(6), 1244–1256 (2014)
11.25 J.A. Crespo, Y. Li, J. Ruiz-Castillo: The measurement of the effect on citation inequality of differences in citation practices across scientific fields, PLOS ONE **8**(3), e58727 (2013)
11.26 R. Fairclough, M. Thelwall: More precise methods for national research citation impact comparisons, J. Informetr. **9**(4), 895–906 (2015)
11.27 M. Thelwall, P. Sud: National, disciplinary and temporal variations in the extent to which articles with more authors have more impact: Evidence from a geometric field normalised citation indicator, J. Informetr. **10**(1), 48–61 (2016)
11.28 M. Thelwall: Three practical field normalised alternative indicator formulae for research evaluation, J. Informetr. **11**(1), 128–151 (2017)
11.29 L. Bornmann, H.D. Daniel: Universality of citation distributions – A validation of Radicchi et al.'s relative indicator $c_f = c/c_0$ at the micro level using data from chemistry, J. Am. Soc. Inf. Sci. Technol. **60**(8), 1664–1670 (2009)
11.30 G. Vaccario, M. Medo, N. Wider, M.S. Mariani: Quantifying and suppressing ranking bias in a large citation network, J. Informetr. **11**(3), 766–782 (2017)
11.31 Z. Zhang, Y. Cheng, N.C. Liu: Comparison of the effect of mean-based method and z-score for field normalization of citations at the level of Web of Science subject categories, Scientometrics **101**(3), 1679–1693 (2014)
11.32 F. Radicchi, C. Castellano: A reverse engineering approach to the suppression of citation biases reveals universal properties of citation distributions, PLOS ONE **7**(3), e33833 (2012)
11.33 Y. Gingras, V. Larivière: There are neither "king" nor "crown" in scientometrics: Comments on a supposed "alternative" method of normalization, J. Informetr. **5**(1), 226–227 (2011)
11.34 H.F. Moed: CWTS crown indicator measures citation impact of a research group's publication oeuvre, J. Informetr. **4**(3), 436–438 (2010)
11.35 T. Opthof, L. Leydesdorff: Caveats for the journal and field normalizations in the CWTS ("Leiden") evaluations of research performance, J. Informetr. **4**(3), 423–430 (2010)
11.36 A.F.J. van Raan, T.N. van Leeuwen, M.S. Visser, N.J. van Eck, L. Waltman: Rivals for the crown: Reply to Opthof and Leydesdorff, J. Informetr. **4**(3), 431–435 (2010)
11.37 P. Vinkler: The case of scientometricians with the "absolute relative" impact indicator, J. Informetr. **6**(2), 254–264 (2012)
11.38 W. Glänzel, B. Thijs, A. Schubert, K. Debackere: Subfield-specific normalized relative indicators and a new generation of relational charts: Methodological foundations illustrated on the assessment of institutional research performance, Scientometrics **78**(1), 165–188 (2009)
11.39 H.F. Moed, R.E. De Bruin, T.N. van Leeuwen: New bibliometric tools for the assessment of national research performance: Database description, overview of indicators and first applications, Scientometrics **33**(3), 381–422 (1995)
11.40 A.F.J. van Raan: Measuring science: Capita selecta of current main issues. In: *Handbook of Quantitative Science and Technology Research*, ed. by H.F. Moed, W. Glänzel, U. Schmoch (Springer, Dordrecht 2005) pp. 19–50
11.41 V. Larivière, Y. Gingras: Averages of ratios vs. ratios of averages: An empirical analysis of four levels of aggregation, J. Informetr. **5**(3), 392–399 (2011)
11.42 L. Waltman, N.J. van Eck, T.N. van Leeuwen, M.S. Visser, A.F.J. van Raan: Towards a new crown indicator: An empirical analysis, Scientometrics **87**(3), 467–481 (2011)
11.43 N. Herranz, J. Ruiz-Castillo: Sub-field normalization in the multiplicative case: Average-based citation indicators, J. Informetr. **6**(4), 543–556 (2012)
11.44 L. Waltman, N.J. van Eck: Field-normalized citation impact indicators and the choice of an appropriate counting method, J. Informetr. **9**(4), 872–894 (2015)
11.45 W.R.J. Tijssen, M.S. Visser, T.N. van Leeuwen: Benchmarking international scientific excellence: Are highly cited research papers an appropriate frame of reference?, Scientometrics **54**(3), 381–397 (2002)
11.46 T.N. van Leeuwen, M.S. Visser, H.F. Moed, T.J. Nederhof, A.F.J. van Raan: The Holy Grail of science policy: Exploring and combining bibliometric tools in search of scientific excellence, Scientometrics **57**(2), 257–280 (2003)
11.47 L. Leydesdorff, L. Bornmann, R. Mutz, T. Opthof: Turning the tables on citation analysis one more time: Principles for comparing sets of documents, J. Am. Soc. Inf. Sci. Technol. **62**(7), 1370–1381 (2011)
11.48 A.I. Pudovkin, E. Garfield: Percentile rank and author superiority indexes for evaluating individual journal articles and the author's overall citation performance, COLLNET J. Sci. Inf. Manag. **3**(2), 3–10 (2009)
11.49 L. Waltman, M. Schreiber: On the calculation of percentile-based bibliometric indicators, J. Am. Soc. Inf. Sci. Technol. **64**(2), 372–379 (2013)
11.50 M. Schreiber: How much do different ways of calculating percentiles influence the derived performance indicators? A case study, Scientometrics **97**(3), 821–829 (2013)
11.51 P. Albarrán, I. Ortuño, J. Ruiz-Castillo: The measurement of low-and high-impact in citation distributions: Technical results, J. Informetr. **5**(1), 48–63 (2011)
11.52 P. Albarrán, I. Ortuño, J. Ruiz-Castillo: High- and low-impact citation measures: Empirical applications, J. Informetr. **5**(1), 122–145 (2011)
11.53 W. Glänzel: High-end performance or outlier? Evaluating the tail of scientometric distributions, Scientometrics **97**(1), 13–23 (2013)

11.54 W. Glänzel, B. Thijs, K. Debackere: The application of citation-based performance classes to the disciplinary and multidisciplinary assessment in national comparison and institutional research assessment, Scientometrics **101**(2), 939–952 (2014)

11.55 W. Glänzel, A. Schubert: Characteristic scores and scales in assessing citation impact, J. Inf. Sci. **14**(2), 123–127 (1988)

11.56 G.-A. Vîiu: Disaggregated research evaluation through median-based characteristic scores and scales: A comparison with the mean-based approach, J. Informetr. **11**(3), 748–765 (2017)

11.57 L. Bornmann, R. Haunschild: How to normalize Twitter counts? A first attempt based on journals in the Twitter index, Scientometrics **107**(3), 1405–1422 (2016)

11.58 A. Schubert, T. Braun: Reference standards for citation based assessments, Scientometrics **26**(1), 21–35 (1993)

11.59 A. Schubert, T. Braun: Cross-field normalization of scientometric indicators, Scientometrics **36**(3), 311–324 (1996)

11.60 C. Colliander: A novel approach to citation normalization: A similarity-based method for creating reference sets, J. Assoc. Inf. Sci. Technol. **66**(3), 489–500 (2015)

11.61 B.I. Hutchins, X. Yuan, J.M. Anderson, G.M. Santangelo: Relative citation ratio (RCR): A new metric that uses citation rates to measure influence at the article level, PLOS Biol **14**(9), e1002541 (2016)

11.62 A.C.J.W. Janssens, M. Goodman, K.R. Powell, M. Gwinn: A critical evaluation of the algorithm behind the relative citation ratio (RCR), PLOS Biol **15**(10), e2002536 (2017)

11.63 B.I. Hutchins, T.A. Hoppe, R.A. Meseroll, J.M. Anderson, G.M. Santangelo: Additional support for RCR: A validated article-level measure of scientific influence, PLOS Biol **15**(10), e2003552 (2017)

11.64 P. Dorta-González, M.I. Dorta-González, D.R. Santos-Peñate, R. Suárez-Vega: Journal topic citation potential and between-field comparisons: The topic normalized impact factor, J. Informetr. **8**(2), 406–418 (2014)

11.65 M. Zitt, H. Small: Modifying the journal impact factor by fractional citation weighting: The audience factor, J. Am. Soc. Inf. Sci. Technol. **59**(11), 1856–1860 (2008)

11.66 L. Leydesdorff, T. Opthof: Scopus's source normalized impact per paper (SNIP) versus a journal impact factor based on fractional counting of citations, J. Am. Soc. Inf. Sci. Technol. **61**(11), 2365–2369 (2010)

11.67 H.F. Moed: Measuring contextual citation impact of scientific journals, J. Informetr. **4**(3), 265–277 (2010)

11.68 W. Glänzel, A. Schubert, B. Thijs, K. Debackere: A priori vs. a posteriori normalisation of citation indicators. The case of journal ranking, Scientometrics **87**(2), 415–424 (2011)

11.69 M. Zitt: Citing-side normalization of journal impact: A robust variant of the audience factor, J. Informetr. **4**(3), 392–406 (2010)

11.70 L. Leydesdorff, L. Bornmann: How fractional counting of citations affects the impact factor: Normalization in terms of differences in citation potentials among fields of science, J. Am. Soc. Inf. Sci. Technol. **62**(2), 217–229 (2011)

11.71 L. Leydesdorff, P. Zhou, L. Bornmann: How can journal impact factors be normalized across fields of science? An assessment in terms of percentile ranks and fractional counts, J. Am. Soc. Inf. Sci. Technol. **64**(1), 96–107 (2013)

11.72 L. Waltman, N.J. van Eck, T.N. van Leeuwen, M.S. Visser: Some modifications to the SNIP journal impact indicator, J. Informetr. **7**(2), 272–285 (2013)

11.73 M. Kosmulski: Successful papers: A new idea in evaluation of scientific output, J. Informetr. **5**(3), 481–485 (2011)

11.74 F. Franceschini, M. Galetto, D. Maisano, L. Mastrogiacomo: The *success*-index: An alternative approach to the *h*-index for evaluating an individual's research output, Scientometrics **92**(3), 621–641 (2012)

11.75 F. Franceschini, D. Maisano: Sub-field normalization of the IEEE scientific journals based on their connection with Technical Societies, J. Informetr. **8**(3), 508–533 (2014)

11.76 J. Nicolaisen, T.F. Frandsen: The reference return ratio, J. Informetr. **2**(2), 128–135 (2008)

11.77 L. Bornmann, R. Haunschild: Citation score normalized by cited references (CSNCR): The introduction of a new citation impact indicator, J. Informetr. **10**(3), 875–887 (2016)

11.78 G. Pinski, F. Narin: Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics, Inf. Process. Manag. **12**(5), 297–312 (1976)

11.79 S. Brin, L. Page: The anatomy of a large-scale hypertextual web search engine, Comput. Netw. ISDN Syst. **30**(1), 107–117 (1998)

11.80 C.T. Bergstrom: Eigenfactor: Measuring the value and prestige of scholarly journals, College Res. Libr. News **68**(5), 314–316 (2007)

11.81 J.D. West, T.C. Bergstrom, C.T. Bergstrom: The eigenfactor metrics: A network approach to assessing scholarly journals, College Res. Libr. **71**(3), 236–244 (2010)

11.82 B. González-Pereira, V.P. Guerrero-Bote, F. Moya-Anegón: A new approach to the metric of journals' scientific prestige: The SJR indicator, J. Informetr. **4**(3), 379–391 (2010)

11.83 V.P. Guerrero-Bote, F. Moya-Anegón: A further step forward in measuring journals' scientific prestige: The SJR2 indicator, J. Informetr. **6**(4), 674–688 (2012)

11.84 L. Waltman, E. Yan: PageRank-related methods for analyzing citation networks. In: *Measuring Scholarly Impact: Methods and Practice*, ed. by Y. Ding, R. Rousseau, D. Wolfram (Springer, Cham 2014) pp. 83–100

11.85 E. Fragkiadaki, G. Evangelidis: Review of the indirect citations paradigm: Theory and practice of

the assessment of papers, authors and journals, Scientometrics **99**(2), 261–288 (2014)
11.86 L. Waltman, N.J. van Eck: The relation between eigenfactor, audience factor, and influence weight, J. Am. Soc. Inf. Sci. Technol. **61**(7), 1476–1486 (2010)
11.87 G. Abramo, C.A. D'Angelo: Ranking research institutions by the number of highly-cited articles per scientist, J. Informetr. **9**(4), 915–923 (2015)
11.88 D. Aksnes, G. Sivertsen, T.N. van Leeuwen, K.K. Wendt: Measuring the productivity of national R&D systems: Challenges in cross-national comparisons of R&D input and publication output indicators, Sci. Pub. Policy **44**(2), 246–258 (2017)
11.89 T. Koski, E. Sandström, U. Sandström: Towards field-adjusted production: Estimating research productivity from a zero-truncated distribution, J. Informetr. **10**(4), 1143–1152 (2016)
11.90 J.E. Hirsch: An index to quantify an individual's scientific research output, Proc. Natl. Acad. Sci. USA **102**(46), 16569–16572 (2005)
11.91 P.D. Batista, M.G. Campiteli, O. Kinouchi, A.S. Martinez: Is it possible to compare researchers with different scientific interests?, Scientometrics **68**(1), 179–189 (2006)
11.92 A.W. Harzing, S. Alakangas, D. Adams: hIa: An individual annual *h*-index to accommodate disciplinary and career length differences, Scientometrics **99**(3), 811–821 (2014)
11.93 J.E. Iglesias, C. Pecharromán: Scaling the *h*-index for different scientific ISI fields, Scientometrics **73**(3), 303–320 (2007)
11.94 J. Kaur, F. Radicchi, F. Menczer: Universality of scholarly impact metrics, J. Informetr. **7**(4), 924–932 (2013)
11.95 F. Radicchi, S. Fortunato, C. Castellano: Universality of citation distributions: Toward an objective measure of scientific impact, Proc. Natl. Acad. Sci. USA **105**(45), 17268–17272 (2008)
11.96 J. Kaur, E. Ferrara, F. Menczer, A. Flammini, F. Radicchi: Quality versus quantity in scientific impact, J. Informetr. **9**(4), 800–808 (2015)
11.97 G. Ruocco, C. Daraio: An empirical approach to compare the performance of heterogeneous academic fields, Scientometrics **97**(3), 601–625 (2013)
11.98 L. Waltman: Special section on size-independent indicators in citation analysis, J. Informetr. **10**(2), 645 (2016)
11.99 G. Abramo, C.A. D'Angelo: A farewell to the MNCS and like size-independent indicators, J. Informetr. **10**(2), 646–651 (2016)
11.100 G. Abramo, C.A. D'Angelo: A comparison of university performance scores and ranks by MNCS and FSS, J. Informetr. **10**(4), 889–901 (2016)
11.101 L. Smolinsky: Expected number of citations and the crown indicator, J. Informetr. **10**(1), 43–47 (2016)
11.102 A. Perianes-Rodriguez, J. Ruiz-Castillo: Multiplicative versus fractional counting methods for co-authored publications. The case of the 500 universities in the Leiden Ranking, J. Informetr. **9**(4), 974–989 (2015)
11.103 P. Albarrán, J.A. Crespo, I. Ortuño, J. Ruiz-Castillo: The skewness of science in 219 sub-fields and a number of aggregates, Scientometrics **88**(2), 385–397 (2011)
11.104 L. Waltman, N.J. van Eck, A.F.J. van Raan: Universality of citation distributions revisited, J. Am. Soc. Inf. Sci. Technol. **63**(1), 72–77 (2012)
11.105 F. Radicchi, C. Castellano: Testing the fairness of citation indicators for comparison across scientific domains: The case of fractional citation counts, J. Informetr. **6**(1), 121–130 (2012)
11.106 L. Leydesdorff, F. Radicchi, L. Bornmann, C. Castellano, W. De Nooy: Field-normalized impact factors (IFs): A comparison of rescaling and fractionally counted IFs, J. Am. Soc. Inf. Sci. Technol. **64**(11), 2299–2309 (2013)
11.107 Y. Li, F. Radicchi, C. Castellano, J. Ruiz-Castillo: Quantitative evaluation of alternative field normalization procedures, J. Informetr. **7**(3), 746–755 (2013)
11.108 L. Waltman, N.J. van Eck: A systematic empirical comparison of different approaches for normalizing citation impact indicators, J. Informetr. **7**(4), 833–849 (2013)
11.109 D. Sirtes: Finding the Easter eggs hidden by oneself: Why Radicchi and Castellano's (2012) fairness test for citation indicators is not fair, J. Informetr. **6**(3), 448–450 (2012)
11.110 F. Radicchi, C. Castellano: Why Sirtes's claims (Sirtes, 2012) do not square with reality, J. Informetr. **6**(4), 615–618 (2012)
11.111 N.J. van Eck, L. Waltman, A.F.J. van Raan, R.J.M. Klautz, W.C. Peul: Citation analysis may severely underestimate the impact of clinical research as compared to basic research, PLOS ONE **8**(4), e62395 (2013)
11.112 L. Leydesdorff, L. Bornmann: The operationalization of "fields" as WoS subject categories (WCs) in evaluative bibliometrics: The cases of "library and information science" and "science & technology studies", J. Assoc. Inf. Sci. Technol. **67**(3), 707–714 (2016)
11.113 Y. Li, J. Ruiz-Castillo: The comparison of normalization procedures based on different classification systems, J. Informetr. **7**(4), 945–958 (2013)
11.114 H.F. Moed: *Citation Analysis in Research Evaluation* (Springer, Dordrecht 2005)
11.115 E.J. Rinia, T.N. van Leeuwen, H.G. van Vuren, A.F.J. van Raan: Comparative analysis of a set of bibliometric indicators and central peer review criteria: Evaluation of condensed matter physics in the Netherlands, Res. Policy **27**(1), 95–107 (1998)
11.116 J. Adams, K. Gurney, L. Jackson: Calibrating the zoom – A test of Zitt's hypothesis, Scientometrics **75**(1), 81–95 (2008)
11.117 L. Bornmann, R. Haunschild: Relative citation ratio (RCR): An empirical attempt to study a new field-normalized bibliometric indicator, J. Assoc. Inf. Sci. Technol. **68**(4), 1064–1067 (2017)
11.118 L. Bornmann, W. Marx: Methods for the generation of normalized citation impact scores in bibliometrics: Which method best reflects the

judgements of experts?, J. Informetr. **9**(2), 408–418 (2015)
11.119 L. Waltman, N.J. van Eck: Source normalized indicators of citation impact: An overview of different approaches and an empirical comparison, Scientometrics **96**(3), 699–716 (2013)
11.120 C. Colliander, P. Ahlgren: The effects and their stability of field normalization baseline on relative performance with respect to citation impact: A case study of 20 natural science departments, J. Informetr. **5**(1), 101–113 (2011)
11.121 A. Perianes-Rodriguez, J. Ruiz-Castillo: A comparison of two ways of evaluating research units working in different scientific fields, Scientometrics **106**(2), 539–561 (2016)
11.122 G. Abramo, C.A. D'Angelo: Evaluating university research: Same performance indicator, different rankings, J. Informetr. **9**(3), 514–525 (2015)
11.123 R.N. Kostoff: Citation analysis of research performer quality, Scientometrics **53**(1), 49–71 (2002)
11.124 R.N. Kostoff, W.L. Martinez: Is citation normalization realistic?, J. Inf. Sci. **31**(1), 57–61 (2005)
11.125 L. Waltman, N.J. van Eck: The need for contextualized scientometric analysis: An opinion paper. In: *Proc. 21st Int. Conf. Sci. Technol. Indic*, ed. by I. Rafols, J. Molas-Gallart, E. Castro-Martínez, R. Woolley (2016) pp. 541–549

Ludo Waltman

Centre for Science and Technology Studies (CWTS)
Leiden University
Leiden, The Netherlands
waltmanlr@cwts.leidenuniv.nl

Ludo Waltman is Senior Researcher at the Centre for Science and Technology Studies at Leiden University in the Netherlands, where he leads the Quantitative Science Studies research group. Ludo's research interests focus on scientometrics and its use in research management and research evaluation. Ludo is particularly interested in scientometric indicators and scientometric visualizations. He is co-developer of VOSviewer and CitNetExplorer, two software tools for scientometric visualization. Ludo is Editor-in-Chief of *Journal of Informetrics*.

Nees Jan van Eck

Centre for Science and Technology Studies (CWTS)
Leiden University
Leiden, The Netherlands
ecknjpvan@cwts.leidenuniv.nl

Nees Jan van Eck is Senior Researcher at the Centre for Science and Technology Studies (CWTS) at Leiden University in the Netherlands. He is conducting research in the field of bibliometrics and scientometrics. Nees Jan is co-developer of VOSviewer and CitNetExplorer, two well-known software tools for visualizing bibliometric data. His current research interests focus on the study of bibliometric data sources and the analysis of full-text scientific publications.

12. All Along the *h*-Index-Related Literature: A Guided Tour

András Schubert, Gábor Schubert

In this chapter, a survey of the literature related to the *h*-index (referred to as *h*-related literature) between 2005 and 2016 is presented. In the first section, the basic definitions and a brief historical account are given. After providing an overview of the typology of the *h*-related publications and some earlier reviews, the more than 3000 *h*-related publications collected from four databases (Web of Science, Scopus, Google Scholar and Microsoft Academic) are analyzed by bibliometric methods. Document types, publication sources, subject categories, geographical distributions, authors and institutions, citations and references are listed and mapped. Several examples of applications of the *h*-index, within and outside the area of scientometrics, are presented, with particular attention to the possibilities for using the *h*-related indices as a network measure. Among the mathematical models used to explain and interpret the index and its relatives, Hirsch's model, the Lotkaian framework, models based on extreme value theory and on fuzzy integrals, and axiomatic approaches are demonstrated.

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_12

12.1 *h*-Index Basics

In 2005, the seminal paper by *Jorge E. Hirsch* introduced a new approach, the h-index, as a way "to quantify an individual's scientific research output" [12.1]. Since then, the h-index has gained great notoriety, as a flood of papers have appeared in efforts to explain, apply, modify, improve, extend or otherwise exploit the concept. This chapter aims to provide am overview of the h-related literature during the period from 2005 to 2016.

12.1.1 Definitions

To begin, we present a few authoritative or commonly found definitions:

> A scientist has index h if h of his or her N_p papers have at least h citations each, and the other $(N_p - h)$ papers have no more than h citations each. [12.1]

> A scientist has h-index equal to H if the top H of his/her N publications from the ranked list have at least H citations each. [12.2]

> $h = \max_i\{C_i \geq i\} i = 1, 2, \ldots, P \wedge C_i \geq C_i + 1$, where P is the number of papers of an author and C_i is the number of citations of the i-th paper, in decreasing order. [12.3]

> h-index $= \max_i \min(f(i), i)$, where $f(i)$ is the number of citations to the i-th publication ordered by decreasing number of citations. [12.4]

A graphical definition of the h-index from a plot of decreasing citations for numbered papers is shown in Fig. 12.1.

12.1.2 The Prehistory of the Index

In a letter to the journal *Nature* in 2005 [12.5], *Anthony W.F. Edwards*, retired Professor of Biometry at Cambridge, shared a memory of his late colleague, Sir Harold Jeffreys, who used a measure similar to Hirsch's proposed index

> for recording his cycling prowess, n being the highest number of days on which he had cycled n or more miles. I think he told me, some 35 years ago, that his n was 70 and that he first had the idea from his fellow cyclist, the astrophysicist Arthur Eddington.

Attempts to give credit to these pioneers by naming the index after Jeffreys or Eddington did not gain wide popularity. Nevertheless, some instances can be found. For example, a post by Leviathan, *What is your Eddington Number?*, appeared on the road.cc bike forum in 2015 [12.6]. And in an article entitled *Calculate your Eddington Number!* on the swinny.net site [12.7], John Swindells presents statistics on cyclists using the Strava app.

Trivia

The above-mentioned Eddington number is not to be confused with another Eddington number, N_{Edd}, which represents the number of protons in the observable universe. In the 1938 Tarner Lecture at Trinity College, Cambridge, Eddington proclaimed:

> I believe there are 15 747 724 136 275 002 577 605 653 961 181 555 468 044 717 914 527 116 709 366 231 425 076 185 631 031 296 protons in the universe and the same number of electrons.

This large number soon became known as the *Eddington number* [12.8].

12.1.3 The Name of the Index

The letter h of the index is generally regarded as a reference to the initial letter of the family name of its creator.

In an interview with *Gualberto Buela-Casal* [12.9], *Hirsch* offered another explanation:

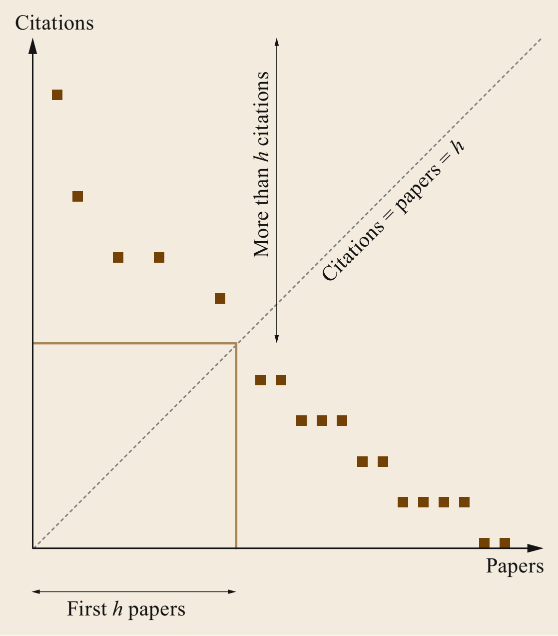


Fig. 12.1 Graphical definition of the h-index (after [12.4])

> I originally thought of calling it 'x index' because it is obtained by determining the intersection between the 'number of citations' (y) versus 'paper number' (x) curve and the $y = x$ line, which leads to an x-shaped graph. Then I thought that 'x' could suggest 'x-rated' so I decided to call it 'h' instead because a high h-index suggests 'highly cited' and 'high achievement'.

Even if it was not his intention, however, Hirsch could not avoid the eponymic association: beginning with *Frangopol*'s paper as early as 2005 [12.10], about 5% of the h-index-related literature refers to the index as the Hirsch index.

A logical choice would be to call the originally conceived formula (h-index for the lifetime achievement of individual researchers) the Hirsch index, and all other indices formed in an analogous manner but with a different premise, Hirsch-type indices or, as shorthand, h-indices. The practice of nomenclature, however, is rarely governed by pure logic.

The attributive *Hirsch-type* appears to have been coined by *Braun*, *Glänzel* and *Schubert* [12.11] to denote indices defined by an algorithm similar to that used by Hirsch in defining the h-index, irrespective of the content of the underlying variables. The term was used in about 2% of the h-index-related literature.

Occasionally, and mainly among journalists and bloggers, the index is also called the h-number or h-factor. It is best to avoid this practice in the scholarly literature, since these terms are widely used in quite different contexts. First, the Hirsch length or Hirsch number of a polycyclic group G is the number of infinite factors in its subnormal series [12.12]. Second, the H-factor is a kinetic model for the rate of delignification in kraft pulping [12.13].

Even the name h-index itself is far from unambiguous. Shannon's diversity index (also referred to as the Shannon–Wiener index) was originally denoted by H', but the name H-index is also frequently used [12.14]. This may be particularly misleading, since it is a statistical indicator of frequency distribution as well. Similarly confusing is the use of H to denote another concentration index: the Herfindahl (also known as Herfindahl–Hirschman) index [12.15]. The term H-index may denote the horizontal (H) component of the index describing geomagnetic disturbance fields [12.16]. It is also typical shorthand for the hemolysis index [12.17], and is used in the acoustic analysis of human speech [12.18].

12.1.4 The Advent of the *h*-Index

Hirsch originally submitted his paper to the Physics and Society section of arXiv on August 3, 2005 [12.19]. The fifth and final version was published September 29, 2005. Hirsch's work gained notoriety after it appeared as the subject of an article by *Philip Ball* published in the journal *Nature* [12.20]. According to the Web of Science, through 2016, Ball's paper was cited in 175 publications, among which more than 90% cited Hirsch's original paper as well.

12.2 A General Overview of the Literature on the *h*-Index

The sudden fame and flame enkindled by Hirsch's paper inspired literally hundreds of scientists and scholars to join the debate. The h-index was celebrated and condemned, specialized and generalized, and improved and deteriorated (although the latter not admittedly) in an ever-growing number of publications. By the end of 2016, more than 3000 publications were connected to Hirsch's paper.

Navigating this literature jungle is facilitated by some excellent reviews.

Alonso et al. [12.21] should be mentioned first, not only because it was the first such review and because of its comprehensiveness and clarity, but also because a somewhat updated version can be found on the authors' *h-Index and Variants* website [12.22], with useful links to the reviewed papers and to several other h-index-related pages.

The work by *Norris* and *Oppenheim* [12.23] appeared almost simultaneously with that of Alonso et al., and in large part addressed the same literature base, although the authors state in the abstract that they "drew on a range of material published in 1990 or so sources published since 2005". Without calling in question the validity of this statement, we find that it is not supported later in the paper.

Two recent books contain extensive literature reviews on the h-index and related topics.

The *Handbook of Bibliometric Indicators* by *Todeschini* and *Baccini* [12.3] is an encyclopedic compilation of all kinds of bibliometric and related indicators. The book contains almost 200 articles (basic indicators or groups of indicators) and more than 1000 keywords in alphabetical order, along with a bibliography of almost 2000 references.

An extensive article on the *h*-index can be found on pp. 162–196, with more than 100 references, commensurate with dedicated review papers.

As the first volume of a new Springer series, *Qualitative and Quantitative Analysis of Scientific and Scholarly Communication*, *Vitanov* published a book with a special focus on bibliometric indicators [12.2]. In Chapter 2 (*Commonly Used Indexes for Assessment of Research Production*), 22 pages (pp. 63–84) and 96 references are devoted to the *h*-index and related indicators. There is also a section in Chapter 5 (*Modeling Production/Citation Process*) where theoretical considerations underlying the *h*-index are treated in detail.

With such excellent review literature available, some of which is quite recent, it seems pointless to reiterate in detail all that has already been said.

Summarizing, and somewhat reconsidering, the above sources, the literature of the *h*-index can be grouped into a few broad categories:

- Introducing the index to specific audiences
  (Fictitious example: *A new index every psychiatrist should know*).
- Application of the index to actual samples
  (Fictitious example: *Top Hungarian oenologists according to their h-index ranking*).
- Comparing/correlating/combining the index with other indicators
  (Fictitious example: *How do the salaries of professors depend on their PageRank and h-index?*).
- Revealing deficiencies (real or imagined) of the index
  (Fictitious example: *The h-index is too complex to reflect real research efforts in soft sciences*).
- Modifying (allegedly, *improving*) the index
  (Fictitious example: *The* $h^{==}$ *index eliminates the effects of social inequalities among researchers*).
- Extending, generalizing the index
  (Fictitious example: *An h-type index in horse betting*).
- Analyzing the theoretical background of the index
  (Fictitious example: *A relativistic quantum h-index for connoisseurs*).

For the majority of papers, of course, categorization is ambiguous or uncertain.

An obvious option, the scientometric analysis of the *h*-index literature, seems to having been somewhat neglected so far. Two papers by *Zhang* et al. [12.24, 25] focus on specific methodological questions rather than providing a general overview.

A substantial part of the *h*-index-related papers argue against the *h*-index. The basis of this criticism is multifarious, and in many respects parallels the objections to the impact factor.

The following provides a typology of objections to the impact factor [12.26, 27]:

(1) Disapproving citation-based or, more generally, all scientometric evaluation
(2) Questioning the suitability of the mean as a characteristic indicator of a sample
(3) Criticizing the choice of publication and citation windows
(4) Condemning the neglect of disciplinary differences
(x) Condemning the neglect of differences between journal types (e.g., review journals)
(5) Condemning the neglect of differences between document types
(6) Condemning the neglect of the role of self-citation
(7) Questioning the use of the impact factor for the evaluation of entities other than journals (individuals, institutions, etc.).

Points (1), (3), (4), (5) and (6) can be transferred to the *h*-index without any change.

Point (x) has relevance only if the *h*-index of journals is considered.

Regarding point (2), the statistical properties of the sample mean are well known, so it is easy to decide whether, in a given case, it is statistically suitable. The *h*-index is much more elusive for standard statistical inquiries. A frequent objection is that it depends on both the sample size (the number of publications) and the distribution of citations. This characteristic is, of course, an inherent property of the *h*-index. The statistical interpretation of this property was therefore an important element of *h*-index research.

Regarding point (7), unlike the case with the impact factor, the use of an index analogous to the *h*-index for objects other than for which it was originally conceived does not seem to present any obstacle. (As we can see, it was even successfully applied to the cyclist community.)

Adding point (8) to the above list, we note that critics often condemn the *h*-index for neglecting the problem of multiauthored papers.

And not mentioned in the above list is point (9): both the impact factor and the *h*-index may be disapproved for the imperfections in the underlying database.

The *disadvantages* of the *h*-index asserted in various sources can largely be classified into one or more of the above categories.

For example, the disadvantages listed in the previously mentioned review by *Alonso* et al. [12.21] (see also the *h-index and Variants* website [12.22]) and in the expedient practical guide to the use of

Table 12.1 Categorization of the reported disadvantages of the h-index

Disadvantage (according to *Alonso* et al. [12.21])	Category
There are inter-field differences in typical h values due to differences among fields in productivity and citation practices, so the h-index should not be used to compare scientists from different disciplines.	(4)
The h-index depends on the duration of each scientist's career, because the pool of publications and citations increases over time.	(3)
Highly cited papers are important for the determination of the h-index, but once they are selected as belonging to the top h papers, the number of citations they receive is unimportant.	(2)
Since the h-index is easy to obtain, we run the risk of indiscriminate use, such as relying only on it for the assessment of scientists. Research performance is a complex, multifaceted endeavor that cannot be adequately assessed by means of a single indicator.	(1)
The use of the h-index could provoke changes in the publishing behavior of scientists, such as an artificial increase in the number of self-citations distributed among the documents on the edge of the h-index.	(1), (6)
There are also technical limitations, such as the difficulty in obtaining the complete output of scientists with very common names, or determining whether self-citations should be removed.	(6), (9)
Disadvantage (according to *Franceschini* and *Maisano* [12.28])	**Category**
h does not take into account multiple co-authorship.	(8)
h does not take into account self-citations.	(6)
h is not useful for cross-disciplinary comparisons because citation rates and scholarly productivity vary considerably among disciplines.	(4)
h does not take into account the age of publications.	(3)
h does not consider the publication type.	(5)
The h-index for a scientist can be easily calculated by using public databases like WoS or GS. Unfortunately, their information can be affected by citation errors—for instance caused by homonymous author names, typographical errors in the source papers, or errors due to some nonstandard reference formats.	(9)

the h-index by *Franceschini* and *Maisano* [12.28] are representative of the categories as indicated in Table 12.1.

Evidently, most of the modifications or improvements proposed in hundreds of papers in the h-index-related literature attempt to circumvent, solve or at least partially remedy these problems.

12.3 Compiling h-Index Bibliographies from Various Bibliographic Databases

Collecting the h-index-related literature is largely facilitated by the fact that a vast majority of papers dutifully cite Hirsch's original paper. Indeed, a citation to Hirsch's paper is an unambiguous sign that the content is somehow related to the h-index. Therefore, works citing Hirsch's paper can be considered a good starting point for a bibliography.

In a bibliometric analysis of the h-index-related literature, *Rousseau* et al. [12.29] compiled a sample of 924 papers counted in the "number of citations received" by Hirsch's original paper in the Web of Science (WoS) between 2005 and 2011 (called H-articles by the authors). Our present analysis is an extension of this effort:

- Extending in time until 2016
- Using a cited reference search to find as many stray citations as possible
- Using and comparing several databases as source
- Complementing citation search with keyword search.

The results of keyword search were carefully cleaned from irrelevant items, where the name Hirsch or the term h-index was used in quite different context.

12.3.1 Web of Science

In the Web of Science (WoS), after careful preliminary studies, the following cited reference search was used:

```
Cited Author: hirsch j* OR hirsh J*
Cited Work: arx* OR ind* OR
an-index* OR p-n* OR pnas OR
natl* OR p*0508025 OR a*0508025
Cited Year: 2005 OR 2006
```

Part B | 12.3

WoS Core Collection (comprising Science Citation Index Expanded (SCI-EXPANDED), Social Sciences Citation Index (SSCI) and Arts & Humanities Citation Index (A&HCI)), Conference Proceedings Citation Index – Science (CPCI-S), Conference Proceedings Citation Index – Social Science & Humanities (CPCI-SSH), Book Citation Index – Science (BKCI-S) and Book Citation Index – Social Sciences & Humanities (BKCI-SSH) have been searched.

Hirsch's paper was found to have been cited in 53 variants (see Appendix 12.A) in a total of 2626 citing papers published between 2005 and 2016.

Two strange "co-authored" versions of Hirsch's paper (in the first and the last row of Appendix 12.A) were the results of errors made by the authors of the citing papers (and in part by the editors of the publishing journals). The authors might confuse the bibliographic data of two consecutive items in the reference list, resulting in a kind of hybrid item. In one case, a correct reference to Hirsch's paper was also present.

In order to find the non-citing *h*-index-related papers, the citation-based list was complemented with the relevant results of an advanced search:

```
TS=(h-index* OR h-indice* OR
((hirsch OR hirsh) SAME (index*
OR indice*))) AND PY=(2005-2016)
```

Irrelevant items were filtered out manually. The citation-based list could thus be complemented with 416 items, resulting in a full WoS bibliographic collection of 3042 papers published between 2005 and 2016.

As we can see, $2626/3042 = 86.3\%$ of the compiled bibliography contained references to Hirsch's original paper. This is an unusually high value, which cannot be explained simply by the originality or the importance of the idea. For instance, in the case of the literature regarding the impact factor (a no less original, important and widely discussed indicator), the original paper by *Garfield* [12.30] collected less than half the number of citations during its full lifetime than did Hirsch's paper, while the total size of the impact factor-related publications was more than four times that of the *h*-index-related literature. The percentage of impact factor-related papers containing reference to Garfield's paper is about 10%, which by and large has remained constant over the last few decades. How strongly a concept is connected to a specific paper is apparently dependent on a multitude of factors that fall outside the realm of bibliometrics.

12.3.2 Scopus

The advanced search

```
REF(hirsch AND j* AND (2005 OR
2006) AND ("index to quantify" OR
0508025))
```

returned 3151 document results published between 2005 and 2016. In addition to 2999 citations in the standard format, 152 stray citations were received by a long list of improper variants. Here we list just a few examples:

Hirsch, J.E., (2005) Arxhiv, 5, pp. 1–5
Hirsch, J.E., (2005) arXiv.org E-Print Archive
Hirsch, J.E., (2005) arXiv: physics/0508025v5 [physics.soc-ph], 29
Hirsch, J.E., (2005) Nature, 436, p. 90
Hirsch, J.E., (2005) Nature, 444, pp. 1003–1004
Hirsch, J.E., (2005) PAMS, 202 (46), pp. 16569–16572
Hirsch, J.E., (2005) Paper presented at the Proc Natl Acad Sci USA
Hirsch, J.E., (2005) PNAS 2005, p. 165691657
Hirsch, J.E., (2005) PNAS, 102 (46), pp. 6569–16572
Hirsch, J.E., (2005) Proc
Hirsch, J.E., (2005) Proc Nat Acad Sci, 102, p. 16572
Hirsch, J.E., (2005) Proc Natl Acad of Sci USA, 572 (46), pp. 516–569
Hirsch, J.E., (2005) Proc Natl Acad Sci U S A
Hirsch, J.E., (2005) Proc Natl Acad Sci USA, 102 (46), pp. 72–16569
Hirsch, J.E., (2005) Proc. Nat. Acad. Sci. U.S.A., 102 (46), pp. 1–5
Hirsch, J.E., (2005) Proc. Nat. Acad. Sci. U.S.A., 46
Hirsch, J.E., (2005) Proc. Natl Acad. Sci. USA, 102, pp. 569–616
Hirsch, J.E., (2005) Proc. Natl. Acad. Sci. USA, 102 (46), pp. 69–72

A keyword search

```
TITLE-ABS-KEY (h-index OR
hirsch-index OR hirsch-type-ind*)
```

resulted in 458 additional papers after removing duplicates and non-relevant items from the citation search results. The total Scopus bibliography thus contained 3608 items.

The percentage share of documents citing Hirsch's paper in the Scopus bibliography is 87.3%, very similar to the WoS value.

There were 2564 papers included in both the WoS and the Scopus bibliography (84% of the WoS, 71% of the Scopus total). The great majority of the papers were included in both databases. A closer look at the lists reveals some reasons for the differences:

- Scopus covers several non-English (mainly Chinese and Spanish) journals not covered by the WoS.
- There is a difference between the two databases in their handling of the most recent literature: Scopus (understandably) is rather prompt in recording Elsevier journals, but does not contain the reference lists until the paper is in press. These papers are, therefore, irretrievable for citation search.
- There are certain random recording errors in the lists of references, obviously different between the two databases. These are not only stray references, but in our case, the notorious Hirsch reference was sometimes simply missing in one or the other database.

The annual number of documents in the two bibliographies and those contained in each is shown in Fig. 12.2. The dynamic increase in the literature until 2013 seems to have slackened lately, but no real decline has yet been encountered.

12.3.3 Google Scholar

In spite of repeated warnings about its unreliability and manipulability [12.31, 32], even the critics admit that the recent version of Google Scholar (GS) "could become a potentially useful complementary resource" [12.33].

Its much wider coverage is reflected by the fact that it records 6110 documents citing Hirsch's paper between 2005 and 2016 (6278 in total; accessed February 13, 2017). Collection of citing documents was attempted using *Anne-Wil Harzing*'s Publish or Perish software [12.34], but because of the inherent limitations of GS, only the 1000 most highly cited publications could be retrieved. About two-thirds of these were included in the WoS or the Scopus citation-based bibliography. The remaining items were mainly books, documents available only on the internet, and articles in less internationally recognized journals.

In our search for items citing Hirsch's paper, we found nice examples for both the pros and the cons of GS. GS was able to identify papers as citing document even if WoS or Scopus was unable to do so. The paper [12.35] which was found to be the most cited paper ever citing Hirsch's work (6976 citations according to the GS), cited it as "J.E. Hirsch, physics/0508025 (2005)". It was properly assigned to the original paper. In our WoS cited reference search, we were able to include this version in the search profile; thus GS indeed became, in Jacsó's words, a "useful complementary resource". Among the false positives, a striking example is another rather highly cited item: G. Placzek: The scattering of neutrons by systems of heavy nuclei. Physical Review, 86(1) (1952) 377–388, receiving 636 citations according to GS. The source of this obvious error is somewhat—but by no means fully—

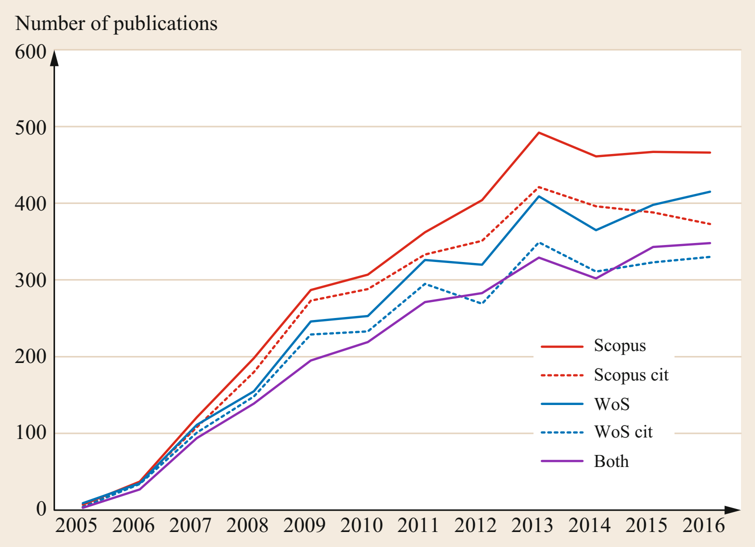


Fig. 12.2 The annual number of h-index-related documents (total and citation-based) in the Web of Science and the Scopus bibliographies and of documents contained in both databases

explained by the fact the PDF link does not point to Placzek's classic paper but to an unpublished manuscript: M. Cardona, W. Marx: Georg(e) Placzek: a bibliometric study of his scientific production and its impact. arXiv:physics/0601113 with references to both Placzek's and Hirsch's papers.

12.3.4 Microsoft Academic

After several years of seeming lifelessness, Microsoft Academic (MA) has lately been revived [12.36, 37] to become a strong competitor of GS.

A citation search for Hirsch's paper for the period 2005–2016 resulted in 3205 citing documents. Unfortunately, it is not possible to produce a complete (or even an incomplete but extensive) list of the citing items (not even with the Publish or Perish software).

We compared the 120 most cited items of the 3205 citing documents with the WoS, Scopus and GS lists. There were 12 items in the MA list not included in any other list: seven were published only on the internet, and three were conference materials. Two papers [12.38, 39], although included in the MA list, did not appear to contain any reference to Hirsch's paper.

12.4 A Bibliometric Overview of the *h*-Index Literature

12.4.1 Document Types

Although some of the document type categories bear different names in the WoS and the Scopus databases, the general features of the composition of the two bibliographies are evident (Fig. 12.3).

Articles
With a share consistently around 70%, the article is by far the leading article type in both the WoS and the Scopus bibliographies. Among articles, the overlap of the two databases is rather high as well: 92% of WoS and 78% of Scopus items are shared.

Conference Materials
In the WoS, conference materials are categorized as either *article; proceedings paper* (those published in regular journals or other serial issues), *proceedings paper* (published in standalone proceedings volumes) or *meeting abstracts* (in journals). These three categories make up the 12% of the WoS bibliography of *h*-index-related papers. The share of conference papers is almost identical in Scopus, but the overlap here is much lower than in the case of articles: 45% of WoS and 32% of Scopus items are shared. The selection in both databases seems to be rather haphazard. Quite frequently, the conference materials, in identical or slightly altered form, were also published as journal articles.

About two-thirds of the *h*-index-related conference papers in the WoS were computer science-related, and 1/4 of them fell into the category of information science & library science.

Editorials, Letters
The relatively high percentage of editorials is no surprise, since the editors of journals, regardless of research area, are those who are most concerned with the standing of their journal or, more generally, the actors (actors, institutes, etc.) in their field. Similarly, the interest (or sometimes anger) of the research community in evaluation-related questions is expressed in letters. This assumption is underscored by the fact that almost 20% of the impact factor-related literature consists of editorials; letters constitute about 4%. For comparison, in the

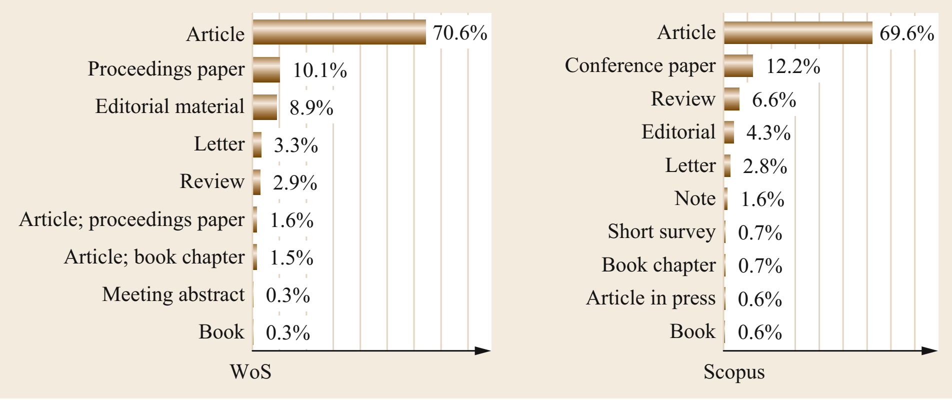


Fig. 12.3 The document type composition of the WoS and the Scopus bibliographies

literature of more neutral topics such as PageRank, percentile or network centrality, editorials are around 0.5% or less, and letters are around 0.1% or less.

Reviews

There appears to be a significant difference in the percentage share of reviews between the WoS and Scopus bibliographies: 2.9 and 6.6%, respectively. However, the difference is by and large illusory. Actually, almost 80% of the reviews found in Scopus can also be found in WoS but are not categorized as reviews. About half of Scopus reviews are articles or editorial materials in WoS.

Books

To our knowledge, no book devoted specifically to the h-index has yet been published.

Two books mentioned earlier [12.2, 3] contain extensive and useful sections on the topic. Neither is included in either of the two databases.

The WoS bibliography includes nine items of document type *book* and 46 *article; book chapter* items; in Scopus, 22 *book* and 27 *book chapter* items are included. The overlap is only around 20%.

Three books, included in both databases, may have special relevance for those interested in the application of the h-index: [12.40–42].

12.4.2 Sources

The 3042 documents of the WoS bibliography are dispersed among 1272 source titles, the 3608 Scopus items among 1584. About 75% of both lists contain only a single item. Almost 90% of the documents in the bibliographies were published in journals.

The top 18 journals with 10 or more papers in the WoS bibliography are shown in Table 12.2.

As the last column of the table shows, the first 10 positions are identical between the two databases. Scopus contains one title among the top producers not covered by WoS: Wuhan Daxue Xuebao (Xinxi Kexue Ban)/Geomatics and Information Science of Wuhan University, with 12 h-index-related publications.

Most of the top titles are as expected. The high activity of some surgery and neurosurgery journals is surprising, but the titles of the papers suggest real interest, even if at times mixed with doubt: "The use of the Hirsch index in benchmarking hepatic surgery research", "Application of the h-Index in Academic Plastic Surgery", "The Hirsch Index; Imaginary Number or High Value Currency?", "Use of the h index in neurosurgery", "Survey of the h index for all of academic neurosurgery: another power-law phenomenon?", etc. The top position of PLoS ONE, while not surprising, is remarkable.

The most productive non-journal sources are the proceedings of the biennial international conferences of the International Society for Scientometrics and Informetrics (ISSI), with 69 papers. The series Lecture Notes in Computer Science (including the subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), covered only by Scopus, published 64 h-index-related papers.

12.4.3 Subject Categories

Papers of both bibliographies can be assigned to subject categories according to the subject classification

Table 12.2 Journals publishing the greatest number of h-index-related papers in the WoS bibliography

WoS rank	Journal title	#	Scopus rank
1	Scientometrics	440	1
2	Journal of Informetrics	232	2
3	Journal of the Association for Information Science and Technology (formerly JASIST)	151	3
4	PLoS ONE	83	4
5	Current Science	36	5
6	Online Information Review	29	6
7	Malaysian Journal of Library & Information Science	21	7
8	Research Evaluation	20	8
9	Information Processing & Management	17	9
10	Journal of Neurosurgery	14	10
11	Profesional de la Informacion	13	13
12	Proceedings of the National Academy of Sciences of the United States of America	12	12
13–16	Journal of Information Science	11	15
13–16	Revista Espanola de Documentacion Cientifica	11	22
13–16	Scientific Reports	11	17
13–16	World Neurosurgery	11	11
17–18	Journal of Surgical Education	10	16
17–18	Nature	10	35

of the source journals. For the WoS, the Essential Science Indicators (ESI) Journal List [12.43] was used, and for Scopus, the Scopus Source List [12.44] was used. The main difference between the two categorization schemes is that ESI assigns each journal to a single category, while Scopus allows for multiple assignments.

The subject composition of the two bibliographies is shown in Fig. 12.4.

Because of the differences in the philosophy of categorization and in the categories themselves, the comparison is difficult and somewhat misleading.

The striking difference in the position of computer science can be attributed to two factors: (1) While in the ESI most of the information science journals are categorized into *social sciences, general*, in Scopus, most of them are assigned *computer science* as a secondary category. (2) A significant portion of the surplus conference papers in Scopus (as compared to the WoS coverage) are computer science-related.

Mathematics also has its strong position in Scopus because of its frequent occurrence as a secondary category.

Table 12.3 Countries publishing the greatest number of *h*-index-related papers in the WoS bibliography

Rank	Country	Code	Papers	Scopus rank
1	United States	US	700	1
2	China	CN	279	2
3	Spain	E	252	3
4	United Kingdom	UK	249	4
5	Germany	G	217	5
6	Belgium	B	168	8
7	Italy	I	166	7
8	Australia	AU	158	6
9	Netherlands	NL	134	12
10	Canada	CA	122	11
11	Brazil	BR	107	9
12	France	F	100	13
13	India	IN	99	10
14	Taiwan	TW	90	14
15	Switzerland	CH	70	15
16	Israel[a]	IS	61	31
17	Poland	PL	54	17
18	Hungary	H	51	19
19	Greece	G	43	18
20	Malaysia	MY	37	20

[a] Including Nablus in the Palestinian territory

12.4.4 World Map of the *h*-Index Literature

The documents in the bibliographies were authored by researchers from about 100 countries. A visual overview of the geographical distribution of authors in the WoS bibliography is given in Fig. 12.5. The detailed national productivity data are given in Table 12.3.

In the Scopus bibliography, Iran takes the 16th position in place of Israel.

12.4.5 Authors and Institutes

More than 6000 authors contributed to the literature on the *h*-index published between 2005 and 2016. The distribution is extremely skewed: about 80% of the authors contributed only a single paper. The most productive authors of the WoS and Scopus bibliographies are listed in Table 12.4.

The two rankings are fully coherent. The top three positions cause no surprise; they are occupied by well-known, prominent researchers of the topic. The subsequent several names are, however, somewhat bewildering. The authors ranked 4–8 do not belong to the mainstream scientometrics community. Their contribution is in the application rather than the development or interpretation of the concept of the *h*-index. As this is a hot topic, such types of publications may also be successful. Their usefulness can even exceed that of many *amelioration* or *clarification* attempts.

Prof. Jean Anderson Eloy and Peter F. Svider are practicing physicians at the Rutgers School of Biomedical and Health Sciences, Department of Otolaryngology and Head & Neck Surgery, Newark, NJ, USA. Among their 95 joint publications (mainly in otorhinolaryngology journals), there are numerous articles oriented toward science policy and research evaluation. Their most highly cited paper is [12.45].

Prof. Waleed M. Sweileh, Sa'ed H. Zyoud and Samah W. Al-Jabi are pharmacologists-toxicologists at the Faculty of Medicine and Health Sciences, An-Najah National University, Nablus, Palestinian territory. They have published 56 joint papers in 38 different journals. A great share of these involve bibliometric analysis of various pharmacological or health-related issues. Their most-cited paper also belongs to this category [12.46].

The authors of the documents in the bibliographies were affiliated with more than 2000 institutions. Again, more than half of the institutions are represented by only one article in the bibliography. Another rather high percentage is represented by more than one paper, but only by one or a very small number of researchers interested in the topic. To further complicate the situation, some authors (among whom are the two most

Fig. 12.5a,b The geographical distribution of authors of the Web of Science bibliography **(a)** in the world and **(b)** in Europe. (The size of the circles is proportional to the number of authors. Codes of major countries are given in Table 12.3. QGIS version 2.18.3) ►

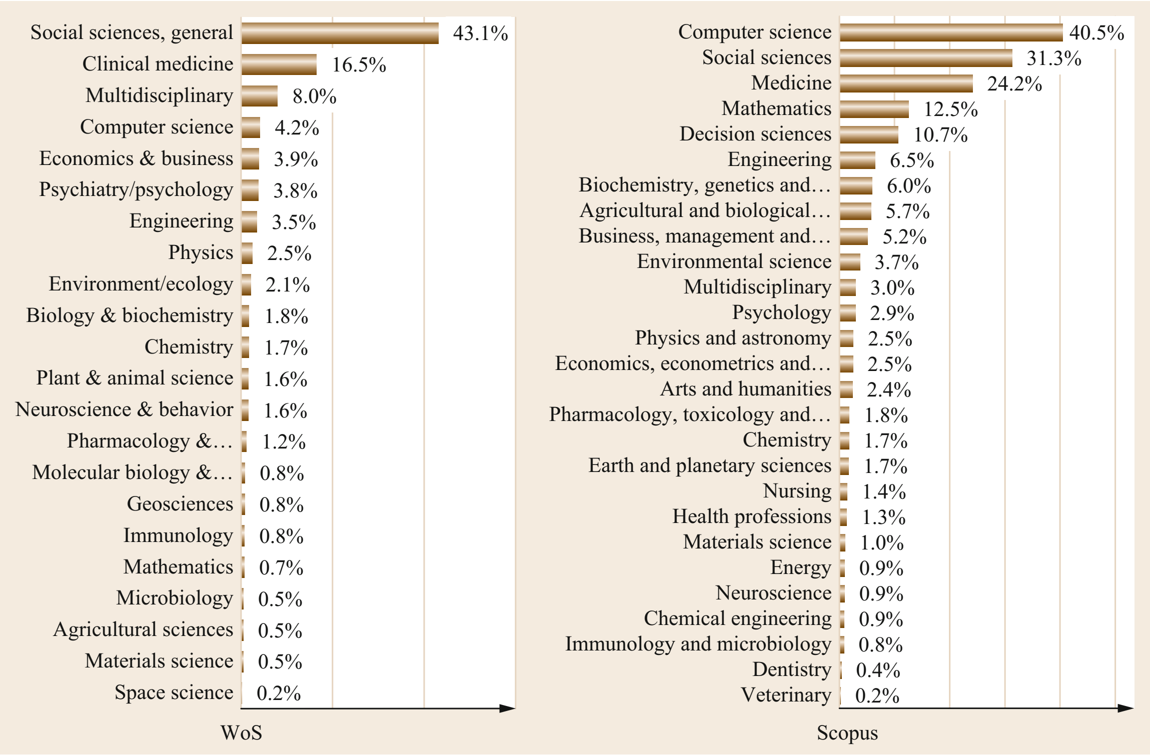


Fig. 12.4 The subject composition of the Web of Science and the Scopus bibliographies

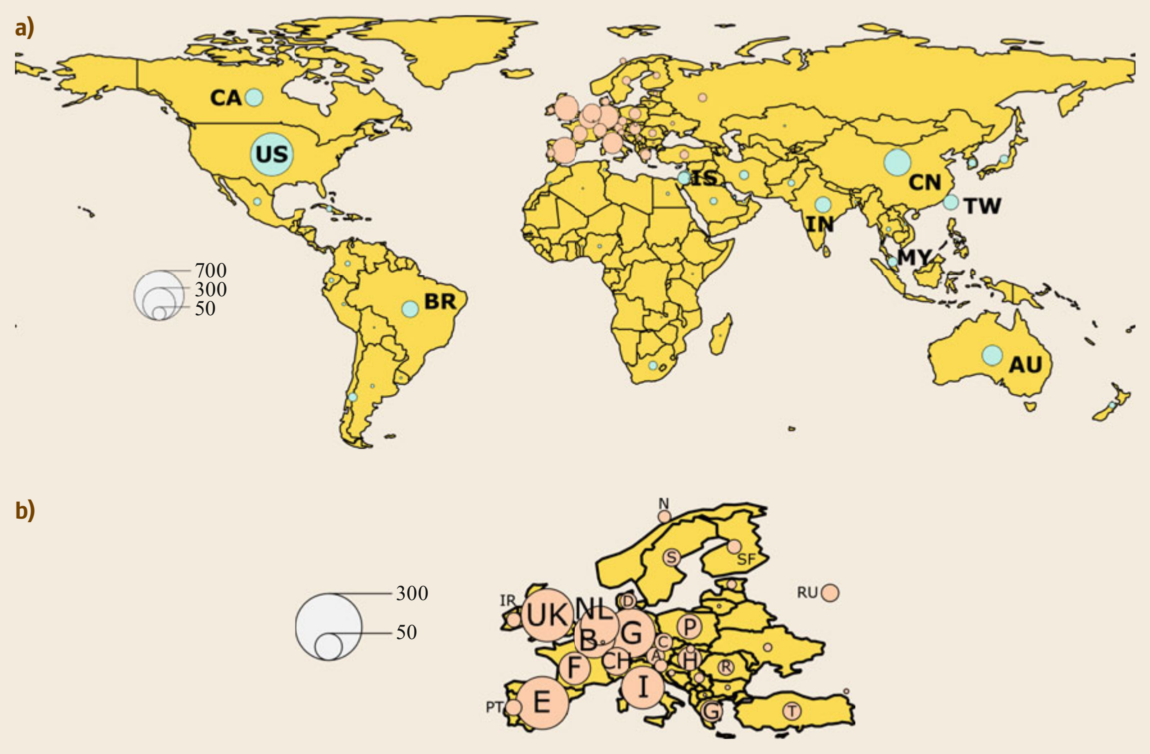


Table 12.4 The most productive authors of the WoS and Scopus bibliographies

WoS rank	Author	Papers	Scopus rank
1	Egghe L.	58	2
2	Rousseau R.	55	1
3	Bornmann L.	51	3
4	Zyoud S.H.	34	4
5	Eloy J.A.	32	7
6	Svider P.F.	29	8
7	Sweileh W.M.	29	5
8	Al-Jabi S.W.	28	6
9	Franceschini F.	28	11
10	Prathap G.	28	9
11	Schreiber M.	26	13–14
12–13	Maisano D.	25	13–14
12–13	Marx W.	25	16
14	Glänzel W.	24	20–21
15–16	Groneberg D.A.	23	10
15–16	Herrera-Viedma E.	23	12
17–18	Jacsó P.	21	23
17–18	Ye F.Y.	21	17–18
19	Gagolewski M.	20	15
20–21	Ho Y.-S.	18	20–21
20–21	Leydesdorff L.	18	24–27
22	Schubert A.	17	24–27
23	Baredes S.	16	24–27
24–29	Abramo G.	15	33–41
24–29	D'Angelo C.A.	15	33–41
24–29	Huang M.-H.	15	32
24–29	Liu J.S.	15	28–31
24–29	Radicchi F.	15	28–31
24–29	Thelwall M.	15	42–52
30–33	Chen D.-Z.	14	33–41
30–33	Cobo M.J.	14	24–27
30–33	Daniel H.-D.	14	28–31
30–33	Sawalha A.F.	14	17–18

productive authors, Leo Egghe and Ronald Rousseau) regularly indicate two or three institutional affiliations in the byline of their publications.

We think, therefore, that for an institutional comparison, it is useful to complement the number of papers with some other indicators. The number of authors indicating the institution as their affiliation is an obvious choice. An h-type indicator can also be calculated: the number, h, of authors each having at least h publications. The size of the h-core is the total number of authors each having at least h publications. It is not necessarily equal to h because of the possible ties. (For a more detailed discussion of the theory and application of the concept of h-core, several excellent papers can be suggested [12.47–49].) The indicators of the 12 most productive institutions in the Scopus bibliography are given in Table 12.5.

Table 12.5 Indicators of the 12 most productive institutions of the Scopus bibliography

Rank	Institute	Papers	Authors	h-Index	h-Core
1	Universiteit Antwerpen	74	6	3	3
2	KU Leuven	71	26	3	5
3	Universiteit Hasselt	61	3	3	3
4	Universidad de Granada	53	62	4	7
5	An-Najah National University	37	16	4	6
6	University of Sydney	36	65	4	7
7	Katholieke Hogeschool Brugge-Oostende	36	1	1	1
8	Rutgers New Jersey Medical School	35	68	5	6
9	Universidade de Sao Paulo	33	58	3	5
10	Indiana University	33	69	5	8
11	ETH Zürich	33	26	4	4
12	Hungarian Academy of Sciences	32	7	3	3

The four indicators together can give a more articulate picture of the author community of the institutions than any single indicator alone. At one extreme, KH Brugge is a "one man band" (Rousseau). Achievements of Antwerpen, Hasselt and the Hungarian Academy of Sciences are produced by isolated small teams. At the other end, in Indiana, Sydney and Granada, 7–8 core authors can attract a considerable number of local coauthors.

12.4.6 Citations

The 3042 publications of the WoS bibliography received a total of 37 451 citations (12.31 citations per paper) from 16 734 citing papers. The h-index of the set of papers is 69.

Twelve papers were cited at least 200 times (Table 12.6).

The h-index is the main topic of seven of these top-cited papers (#2, #3, #6, #8, #9, #10 and #11), has a *supporting role* in #7 and is mentioned only passingly in #1, #4 and #5. In all of these papers, Hirsch's origi-

Table 12.6 The most cited papers of the WoS bibliography

#	Author(s)	Title	Journal	Year	Vol	Iss	Pages	Cites
1	Boccaletti, S; Latora, V; Moreno, Y; Chavez, M; Hwang, DU	Complex networks: Structure and dynamics	Physics Reports—Review Section of Physics Letters	2006	424	4/5	175–308	4001
2	Hirsch, JE	An index to quantify an individual's scientific research output	Proc. Natl. Acad. Sci. USA	2005	102	46	16569–16572	2484
3	Egghe, L	Theory and practise of the g-index	Scientometrics	2006	69	1	131–152	582
4	Chen, HC; Chiang, RHL; Storey, VC	Business intelligence and analytics: From big data to big impact	MIS Quarterly	2012	36	4	1165–1188	368
5	Meho, LI; Yang, K	Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar	JASIST	2007	58	13	2105–2125	348
6	Hirsch, JE	Does the h index have predictive power?	Proc. Natl. Acad. Sci. USA	2007	104	493	19193–19198	316
7	Van Raan, AFJ	Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups	Scientometrics	2006	67	3	491–502	276
8	Radicchi, F; Fortunato, S; Castellano, C	Universality of citation distributions: Toward an objective measure of scientific impact	Proc. Natl. Acad. Sci. USA	2008	105	45	17268–17272	255
9	Jin, BH; Liang, LM; Rousseau, R; Egghe, L	The R- and AR-indices: Complementing the h-index	Chinese Science Bulletin	2007	52	6	855–863	233
10	Braun, T; Glänzel, W; Schubert, A	A Hirsch-type index for journals	Scientometrics	2006	69	1	169–173	231
11	Batista, PD; Campiteli, MG; Kinouchi, O; Martinez, AS	Is it possible to compare researchers with different scientific interests?	Scientometrics	2006	68	1	179–189	214
12	Bar-Ilan, J	Which h-index?—A comparison of WoS, Scopus and Google Scholar	Scientometrics	2008	74	2	257–271	200

nal paper is formally referenced (in #2, Hirsch's paper itself, the arXiv variant).

Paper #1 in Table 12.6 is a real citation superstar, a comprehensive review of the state of the art of network theory in 2006 based on 888 references. Reference [551] is Hirsch's paper cited as "J.E. Hirsch, physics/0508025 (2005)". It is cited together with *Ball*'s paper [12.20] in the context

> recently the power law behavior of the in-degree distribution suggested the introduction of a new indicator (the index h) quantifying the impact of a scientist on the modern scientific community.

It was able to be retrieved as citing document to Hirsch's paper in GS and WoS, but not in MA or Scopus. In MA it was not identified as a linked reference; in Scopus, the references were recorded as "Hirsch JE (2005) 1" without title or source.

The Boccaletti et al. paper was co-cited with Hirsch's paper seven times.

The h-index was used as citation indicator for authors and journals, respectively (Tables 12.7 and 12.8).

Both the author and the journal rankings clearly show strong similarities to and characteristic differences from the productivity rankings (Tables 12.4 and 12.2, respectively). To facilitate comparison, the productivity rank is given in the last column.

12.4.7 References

The 3042 publications of the WoS bibliography contained more than 100 000 references to approximately 50 000 documents. Among the references, 18% were made to 3% of the documents: to about 1500 items

Table 12.7 The *h*-index ranking of authors in the WoS bibliography

Rank	Author	Papers	Cites	Cites/Paper	*h*-Index	Prod rank
1	Bornmann, L.	51	1436	28.2	19	3
2	Eloy, J.A.	32	646	20.2	17	5
3	Svider, P.F.	29	593	20.4	16	6
4	Rousseau, R.	55	865	15.7	15	2
5–8	Baredes, S.	16	450	28.1	12	23
5–8	Daniel, H.-D.	14	928	66.3	12	30–33
5–8	Egghe, L.	58	1505	25.9	12	1
5–8	Radicchi, F.	15	663	44.2	12	24–29
9–11	Glänzel, W.	24	736	30.7	11	14
9–11	Jacsó, P.	21	363	17.3	11	17
9–11	Schreiber, M.	26	453	17.4	11	11
12–13	Schubert, A.	17	678	39.9	10	22
12–13	Waltman, L.	13	448	34.5	10	34–37
14–22	Abramo, G.	15	192	12.8	9	24–29
14–22	Al-Jabi, S.W.	28	221	7.9	9	8
14–22	D'Angelo, C.A.	15	192	12.8	9	24–29
14–22	Ding, Y.	12	341	28.4	9	38–42
14–22	Ho, Y.S.	18	298	16.6	9	20–21
14–22	Leydesdorff, L.	18	483	26.8	9	20–21
14–22	Setzen, M.	12	273	22.8	9	38–42
14–22	Sweileh, W.M.	29	222	7.7	9	7
14–22	Zyoud, S.H.	34	249	7.3	9	4

Table 12.8 The *h*-index ranking of journals in the WoS bibliography

Rank	Journal	Papers	Cites	Cit/paper	*h*-Index	Prod rank
1	Scientometrics	440	6666	15.2	36	1
2	Journal of Informetrics	232	3581	15.4	30	2
3	Journal of the Association for Information Science and Technology (formerly JASIST)	150	3345	22.3	30	3
4	PLoS ONE	83	1046	12.6	16	4
5	Online Information Review	29	330	11.4	11	6
6	Proceedings of the National Academy of Sciences of the USA	12	3032	252.7	10	12
7	Information Processing & Management	17	232	13.6	9	9
8–10	Journal of Information Science	11	203	18.5	8	13–16
8–10	Nature	10	423	42.3	8	17–18
8–10	Research Evaluation	20	156	7.8	8	8
11	Journal of Neurosurgery	14	288	20.6	7	10
12–16	Bioscience	8	97	12.1	6	24–27
12–16	Energy Education Science and Technology Part A–Energy Science and Research	7	217	31.0	6	28–31
12–16	Laryngoscope	9	230	25.6	6	19–23
12–16	Physical Review E	9	274	30.4	6	19–23
12–16	Scientific Reports	11	115	10.5	6	13–16
17–22	Academic Radiology	9	151	16.8	5	19–23
17–22	Archivum Immunologiae et Therapiae Experimentalis	5	199	39.8	5	44–60
17–22	Current Science	36	117	3.3	5	5
17–22	EMBO Reports	8	213	26.6	5	24–27
17–22	Journal of Surgical Education	10	70	7.0	5	17–18
17–22	Otolaryngology–Head and Neck Surgery	9	157	17.4	5	19–23

Table 12.9 Papers receiving the greatest number of citations from the WoS bibliography items

Rank	Reference	Cites in the bibliography	Total cites	Rank in the bibliography by total cites
1	Hirsch JE, 2005, P Natl Acad Sci USA, V102, P16569	2626	2626	2
2	Egghe L, 2006, Scientometrics, V69, P131	567	582	3
3	Hirsch JE, 2007, P Natl Acad Sci USA, V104, P19193	277	316	6
4	Van Raan AFJ, 2006, Scientometrics, V67, P491	239	276	7
5	*Garfield E, 2006, JAMA-J Am Med Assoc, V295, P90*	*229*	*759*	–
6	Jin BH, 2007, Chinese Sci Bull, V52, P855	229	233	9
7	Braun T, 2006, Scientometrics, V69, P169	215	231	10
8	Batista PD, 2006, Scientometrics, V68, P179	203	214	11
9	Alonso S, 2009, J Informetr, V3, P273	184	195	14
10	*Garfield E, 1955, Science, V122, P108*	*180*	*879*	–
11	Ball P, 2005, Nature, V436, P900	175	175	21
12	Bornmann L, 2007, J Am Soc Inf Sci Tec, V58, P1381	173	178	20
13	Bornmann L, 2008, J Am Soc Inf Sci Tec, V59, P830	171	179	19
14	*Seglen PO, 1997, Brit Med J, V314, P498*	*167*	*1054*	–
15	*Garfield E, 1972, Science, V178, P471*	*164*	*1275*	–
16	Meho LI, 2007, J Am Soc Inf Sci Tec, V58, P2105	159	348	5
17	Egghe L, 2006, Scientometrics, V69, P121	155	166	24
18	*Egghe L., 2006, ISSI Newsletter, V2, P8*	*151*	*157*	–
19	Bornmann L, 2005, Scientometrics, V65, P391	147	150	27
20	Bar-Ilan J, 2008, Scientometrics, V74, P257	145	200	12

of the bibliography itself. Not surprisingly, the papers receiving the greatest number of citations from the bibliography items were also among the most highly cited papers of the bibliography (Table 12.6).

The 20 references receiving the greatest number of citations from the WoS bibliography items are listed in Table 12.9. The number of citations received from the bibliography items is supplemented with the total number of citations and the paper's rank in the citation-ranked list of the bibliography (Table 12.6), provided that it was included in it. The overwhelming majority of the citations of the bibliography items came from other bibliography items, indicating the high degree of coherence of the h-index-related literature.

All five papers in the top-20 list not included in the bibliography (they are italicized in Table 12.9) dealt with the impact factor. Three were published well before Hirsch's paper.

The oldest work co-cited with Hirsch's paper is Johannes Kepler's Astronomia Nova from 1609. The citing work is [12.50].

Among the co-cited authors, one may find Newton, Swift, Adam Smith, Hume, Lavoisier, Humboldt, Darwin and even Karl Marx (with his Communist Manifesto) [12.51].

12.5 Application of the h-Index Concept Within and Outside the Realm of Bibliometrics

In his paper [12.1], *Hirsch* definitively set forth the purpose of the h-index: "to quantify an individual's scientific research output". The index, however, soon outgrew the original intent, becoming a useful statistical indicator for variables and conditions quite different from the original settings. The differences include various types, as follows:

1. Citations to publications not of an individual but of groups, institutions, journals, countries, etc. are considered (e. g., h-index for journals in SCImago).
2. Citation and publication windows are limited (e. g., Google Scholar's h5-index: the h-index for articles published within the last 5 complete years.
3. Citation and publication numbers are adjusted (e. g., discount of self-citations, partial credit to co-authors).
4. Variables other than citations to publications (e. g., visits to a web page) are considered.
5. The algorithm for determining the index value is modified (e. g., Kosmulski's $h^{(2)}$-index: the highest natural number such that the author's $h^{(2)}$ most

cited papers received each at least $(h^{(2)})^2$ citations [12.52]).

There is no consensual naming convention for these types of variants. *Boell* and *Wilson* introduced the term *h-like index* for indices using the same rationale (algorithm) as the h-index but which are not based on citations [12.53]. This proposal, however, did not gain wide acceptance. *Harzing* [12.34], for instance, calls Schreiber's multi-authored h-index [12.54] an h-like index, although it is clearly citation-based. In *Vitanov*'s book [12.2], the term *h-like index* is used in a vague sense to refer to indices resembling or related to the h-index.

Our modest proposal is as follows: reserve the name h-index (or Hirsch index) for the index that complies with Hirsch's original intentions (citation-based measure for individuals according to Hirsch's algorithm); call all indices using Hirsch's original algorithm h-type indices (within this category one can distinguish citation-based from non-citation-based indices); and refer to indices as h-related indices whenever the algorithm itself is modified (again, with citation-based and non-citation-based subcategories).

12.5.1 Application of the *h*-Index (in the Strict Sense)

Following Hirsch's example of listing top physicists by their h-index, a vast array of publications have included h-ranked lists of eminent researchers of various fields, as *Meyer* describes: from "the psychologists of Italy [...] to the chemists of Peru, not to forget the library scientists of Hungary" [12.55].

Publication of h-index-based lists of individuals does not necessarily meet with unequivocal positive acceptance. Henry Schaefer and Amy Peterson of the University of Georgia, GA, USA, published an h-index ranking of living chemists [12.56]. The list was first announced in April 2007 [12.57], and was regularly updated until December 2011. In compiling their rankings, Schaefer and Peterson manually combed through data from several databases in an attempt to match papers to authors and ensure the accuracy of their data set. Due to constant fierce criticism, however, they eventually decided to abandon the project [12.58].

In the face of potential critics, the Webometrics website in its January 2017 edition published a list that includes both living and deceased authors of all research fields ranked by their h-index [12.59] on the basis of Google Scholar data. According to this list, the all-time champion of the h-index is Sigmund Freud, with an h-index of 269.

12.5.2 Application of Citation-Based *h*-Type Indices

Levels of Aggregation

At higher-than-individual level, the h index has been used to quantify the research output of research groups [12.60–62], institutions [12.63–67], countries [12.68, 69] and journals [12.11, 70, 71].

A considerable body of literature has been published on the application of the h-index for journals in various disciplinary and geographical areas, often in comparison with the impact factor. Some examples are compiled in Table 12.10.

Direct comparisons with the impact factor generally favor the h-index:

> h-index would be a better citation-based metric for evaluating the quality and contribution of scholarly journals than other metrics such as the impact factor (IF) or the number of cites per paper (cpp). [12.75]

> h-index provides a better alternative for journal rankings than the ISI JIF. It does not suffer from the same statistical limitations as the JIF and is more suitable to measure a journal's wider economic or social impact rather than its impact on an academic audience only. As such we argue it provides a more accurate and more comprehensive measure of journal impact. [12.79]

Table 12.10 Application of the h-index for journals in various disciplinary and geographical areas

Area	Reference
Anesthesia	[12.72]
Biomedicine (Spanish)	[12.73]
Business	[12.74]
Business and management	[12.75]
Chemical engineering	[12.76, 77]
Ecology	[12.78]
Economics and business	[12.79]
Forestry	[12.80]
Horticulture	[12.81]
Latin America	[12.82]
Medical education	[12.83]
Pharmacology	[12.84]
Pharmacology and psychiatry	[12.85]
Psychiatry	[12.86, 87]
Psychology	[12.88]
Reproduction biology	[12.89]
Social work	[12.90]
Soil research	[12.91]
Toxicology	[12.92]

> *h*-index [...] may be more useful than the Journal Impact Factor, as a measure of journal quality, and in providing a basis to rank journals. [12.80]

> despite certain flaws and weaknesses, [the *h*-index] is considerably better than using a journal's impact factor. [12.87]

> *h*-index represents an important complement, and perhaps improvement, to the use of impact factors as a way to assess journal quality. [12.90]

As specific advantages, its robustness and its good correlation with peer evaluations is generally stressed.

From a more balanced perspective:

> *h*-index is a useful supplementary indicator, enrichment for the bibliometrics toolset but it is certainly not suited to substitute advanced indicators (JIF) which have long ago become standard in bibliometric research. [12.89]

A paper with a title that appears contradictory to the superiority of the *h*-index [12.93] actually deals with a quite different question: whether the impact factor of a journal or the *h*-index of its first author is a better predictor of the citation future of a paper.

The *h*-index can be determined for any aggregate of publications, not only for groups of researchers or papers of a journal. Not long after the publication of Hirsch's paper, *Banks* applied the concept to chemical compounds and scientific topics [12.94]. Papers of a selected set of compounds and topics were compiled, and the citation-based *h*-index of the sets was determined. Considering the publication year of the paper (2006), it is not surprising that C_{60}, fullerenes and carbon nanotubes dominated both the chemistry and the physics scenes, challenged only by magnetoresistance and quantum dots in physics.

McIntyre and colleagues put the idea into action by using the topic-based *h*-index as a proxy for the impact of human and domestic animal pathogens [12.95–97]. A strong correlation was found between burden of disease (in log10-transformed disability-adjusted life years (DALYs)) and *h*-index scores. *Escherichia coli* was found to be the highest-impact pathogen in Europe in both humans and domestic animals, followed in humans by the HIV1 and HIV2 and the hepatitis C virus. *McIntyre* later extended the method to salmon pathogens, where sea lice and furunculosis were found to be key diseases [12.98].

The top three pathogens in swine were *Escherichia coli*, porcine reproductive and respiratory syndrome virus, and porcine circovirus type 2, with *h*-indices of 106, 95 and 85, respectively [12.99].

Time Periods

It is common practice to contrast the short-term impact measured by the impact factor with the lifetime impact measured by the *h*-index. Although it is undeniably true if we compare the impact factors of the *Journal Citation Reports* (JCR) back in the twentieth century with Hirsch's original proposal, both indicators have turned out to be much more flexible than that. The present-day JCR contains 5-year impact factors, and Google Scholar offers the *h*5, a similar 5-year *h*-type index. The substantial difference, however, lies in the algorithm: the impact factor is a sample mean, while the *h*-index is determined using the specific algorithm proposed by Hirsch. Consequently, the impact factor is size-independent, while the *h*-index is dependent on the sample size.

12.5.3 Application of Non-Citation-Based *h*-Type Indices

Unweighted Graphs

There is an obvious application of the *h*-index concept independent of any actual representation or implementation: its use as a network indicator. To each node of a graph, the vector of the degrees of its neighboring nodes can be assigned, and after arranging the vector in decreasing order of degrees, the *h*-index of the node can be determined. From the vector of *h*-indices of the nodes of a graph, analogously, the *h*-index of the whole graph can be determined. There is an obvious analogy with the various graph centrality concepts, where the centrality distribution of the nodes determines the centralization of the graph.

The idea was so self-evident that it was almost simultaneously published in two independent versions [12.100, 101]. A more elaborate introduction of the index was published soon after the scientometrics-oriented application in the *Korn* et al. paper [12.100] under the name *lobby index* [12.102]. The lobby index of a node x is the largest integer k such that x has at least k neighbors with a degree of at least k.

The role of the photo-finish camera is taken nowadays by preprint services. The first version of the *Korn* et al. paper [12.102] was posted on arXiv as arXiv:0809.0514v1 [physics.soc-ph] 2 Sep 2008 (the final version as arXiv:0809.0514v6 [physics.soc-ph] 26 Jan 2009), while the *Eppstein–Spiro* paper [12.101] was posted as arXiv:0904.3741v1 [cs.DS] 23 Apr 2009 (first and final version).

The *Eppstein–Spiro* paper [12.101] was later published as a journal article [12.103].

Both versions were cited several times, although in different contexts: the *Eppstein–Spiro* paper mainly in the context of subgraph problems [12.104–107], and the lobby index considered rather as an alternative to the traditional centrality measures. To our knowledge, the two sources have never been cited together—until this chapter.

The application of the *h*-type network centrality measure, whether named lobby index or not, is rather diverse and usually successful. It has been used, for example, in:

- Bibliometric network analysis [12.49]
- Designing information-forwarding algorithms in social networks, where "Lobby Influence outperformed Bubble Rap and Epidemic routing algorithms in terms of message delivery and speed" [12.108]
- Proteomics [12.109]
- Information search on complex networks [12.110]
- Improving multimedia content delivery, where "the use of the Lobby-index score metric results in higher hit ratio percentages compared to the equivalent HITS score cases" [12.111]
- Building a conceptual framework for characterizing the ego in a network [12.112].

The relation of the *h*-type network index to other centrality measures was also studied in detail [12.113].

The application of the *h*-core of clusters in local clustering to the delineation of core documents also deserves mention here [12.49].

Weighted Graphs

The generalization of the network *h*-index (lobby index) to weighted graphs is straightforward. Let us define the strength of a node in a weighted network as the sum of the weights of all its links. Then, the *h*-degree of node n in a weighted network is equal to h if h is the largest natural number such that n has at least h links each with strength at least equal to h [12.114].

The *h*-degree was successfully used in finding basic structural patterns in weighted networks [12.115, 116].

A typical weighted graph in scientometrics is the co-author network, where the weights are the number of occasions a pair of authors had joint publications. *Abassi* defined seven new measures for network nodes, and tested them on co-authorship networks [12.117]:

- Definition 1. The al-index of a node is the average degrees of the top k neighbors of the node that each has at least k links (degree of k).
- Definition 2. The gl-index of a node is the largest number such that the top k neighbors of the node have together at least k^2 links (connections).
- Definition 3. The a-degree of a node is the average weights of the top k links of the node that have minimum weight of k.
- Definition 4. The g-degree of a node is the largest number such that the top k links of the node have together a minimum weight of k^2.
- Definition 5. The Hw-degree of a node is the largest number such that the top k neighbors of the node have each at least nwD of k (the *neighbor-weighted degree* (nwD) is the degree of each neighbor multiplied by the weight of the link between that neighbor and the node).
- Definition 6. The Aw-degree of a node is the average of the nwD of the top k neighbors of the node that have nwD of at least k.
- Definition 7. The Gw-degree of a node is the largest number such that the top k neighbors of the node have together at least a nwD of k^2.

Searching for correlation between the authors' centrality measures and their bibliometric performance measures, the author found that "the average based centrality measures (a-degree and Aw-degree) show better correlation with performance".

A further step in generalization is the directed *h*-degree in directed, weighted networks [12.118]. The in-*h*-degrees and out-*h*-degrees are defined in line with *h*-degrees in undirected graphs [12.114]:

- Definition 1. In a directed weighted network, the in-*h*-degree of node n is equal to h if h is the largest natural number such that n has at least h in-links each with strength at least equal to h.
- Definition 2. In a directed weighted network, the out-*h*-degree of node n is equal to h if h is the largest natural number such that n has at least h out-links each with strength at least equal to h.

The prototype of directed graphs in scientometrics is the citation network. The in-*h*-degree/out-*h*-degree concept was tested on the citation network of 46 library & information science journals. JASIST was found to be the most central journal of the field with regard to both the in- and the out-*h*-degree.

A special case of the *h*-degree was defined as the PartnersHIp Ability Index (PHI, φ) [12.119]. A partnership comes about between two actors through a joint action. An actor is said to have a partnership ability index, φ, if the φ of his/her n partners had at least φ joint actions each, and with the other $(n-\varphi)$ partners, had no more than φ joint actions each. The basic properties of the indicator are as follows:

- The partnership ability is 0 if and only if the actor has no joint action with any other actor.
- The partnership ability is 1:
  (a) if the actor had an arbitrary number of joint actions with the very same partner,

(b) if the actor had an arbitrary number of partners with no more than one joint action with each, or
(c) if the actor had an arbitrary number of joint actions with the very same partner AND an arbitrary number of partners with no more than one joint action with each [12.120].

- In all other cases, the partnership ability is an integer larger than 1.
- The use of the index is exemplified in the paper by a co-authorship example, but it is stressed that

> the possible use of the φ-index is far not restricted to scientometrics. As suggested in the definition, actors in all kind of co-activity network can be characterized this way. The word 'actor' may be, e. g., taken literally [... by] the analysis of the Internet Movie Database (IMDb). A delicate possible area of application might be that of sexual encounter networks widely studied nowadays to understand and control the propagation of sexually transmitted diseases [12.119].

The application of the φ-index was tested at a much larger scale as well: on the co-authorship network of over a million Data Bibliography and Library Project (DBLP) authors [12.121].

Along with some further co-authorship network examples [12.122–124], the φ-index was also applied for characterizing the partnership ability of jazz musicians [12.125].

A similar but somewhat different measure, the collaboration index or communication centrality or c-index, was defined to characterize *collaboration competence* [12.126] or *communication ability* [12.127]. The c-index of node x is the largest integer c such that the node x has at least c neighboring nodes satisfying the condition that the product of each node strength and the strength of the edge linked with node x is not less than c (cf. the definition of h-degree [12.114]).

Other Miscellaneous Applications

Social media is an inexhaustible source of network analysis, and obviously, h-type indices have found their way to this area as well [12.128–131]. One remarkable idea, for example, is measuring the controversial nature of a post with an h-type index of the comments (as "publications") and replies (as "citations")—originally proposed for Slashdot [12.128] but easily extended to other social media.

An h-type index characterizing chemical compounds was already mentioned earlier [12.94]. It was, however, a citation-based index derived from the bibliometric analysis of the literature of the compounds under study. An h-type index concerning the concentration of the chemicals was proposed as a tool for compound prioritization in environmental monitoring [12.132]. In order to make the idea work, the concentration values had to be properly rescaled to make them commensurate with the ranks in the ordered list (actually, normalization to 100 is used in the paper). A general scheme of rescaling the h-index for whatever purpose [12.133], and a special h-type index for distributions of percentage-valued variables [12.134], can also be found in the literature.

Sports is a typical training ground for all kinds of statistical exercises. Apart from the historical example of cycling mentioned earlier, it is surprisingly difficult to find a good sports-related example of h-type indices. The first one, to our knowledge, concerned cricket [12.135].

An analogy was found between the h-index and the rule for fixing the size of the group of athletes a National Olympic Committee (NOC) may select for each individual event at the Olympic Games [12.136].

12.5.4 Application of h-Related Indices

In a recent review of the literature on citation impact indicators [12.137], *Waltman* formulated four recommendations—not specifically for the h-type and h-related indices, but for all citation-based indicators:

- Recommendation 1: Do not introduce new citation impact indicators unless they have a clear added value relative to existing indicators.
- Recommendation 2: Pay more attention to the theoretical foundation of citation impact indicators.
- Recommendation 3: Pay more attention to the way in which citation impact indicators are being used in practice.
- Recommendation 4: Exploit new data sources to obtain more sophisticated measurements of citation impact.

In view of the proliferation of h-related literature in general, and h-related indices in particular, these recommendations seem to be more apt than ever.

In the review by *Alonso* et al. [12.21] and its somewhat updated web version [12.22], 47 different h-related index versions were compiled. In the recent handbook of *Todeschini* and *Baccini* [12.3], 59 such indicators are listed, partly overlapping with the previously mentioned compilation. In Appendix 12.B, a unified alphabetical list of the h-related indices included in these two sources is given. The original sources of these indices can only be partially found in the references of this chapter; as to the rest, the reference lists of the two mentioned review sources may be visited.

As a matter of fact, although most of these indices claim to remedy some real or perceived deficiency or weakness of the original h-index (adjustments for multi-authorship, self-citation, disciplinary differences, etc.), very few of them have ever been used in real-life situations. If they have ever been cited, it is self-citation or inclusion in reviews or other compilations. Frankly, we did not want to add to the number of these idle citations.

Instead, we cite a very important lesson from one of the h-related editorials:

> Is the h-index a more useful metric than the [impact factor]? That depends: more useful for what purpose? The point is not to argue in favor of the h-index but rather to reiterate one of the most fundamental lessons ...: namely, the methods used ... should be foremost informed by the question we are asking. The close corollary of this lesson is that without a clear question, we cannot evaluate the adequacy of the methods, much less their results. [12.138]

The message can be readily transferred to the comparison of the h-index and its variants. The h-index has several features. It is a combined measure of publication output and citation impact. It is not sensitive to outstanding citation rates. It is not sensitive to a massive amount of uncited or poorly cited papers. Its statistical behavior is not as straightforward as, say, an average, but—as we will soon see—can be treated sufficiently with appropriate statistical tools.

It is the question we are asking that determines whether these features are positive or negative, whether the h-index is a proper measure to answer the question. And if the answer is that it is not, then there is still an arsenal of existing bibliometric indicators (see, e. g., the *Handbook of Bibliometric Indicators* [12.3]) from which to find a more suitable one. Even if the question is clear and well defined, the introduction of a new indicator is not justified by its mere superiority to the h-index (what might well be completely unsuitable in certain cases) but by its superiority among all existing and well-established indicators. This is what is painfully lacking from a substantial number of the attempts to introduce new h-related indicators.

Without doubt, the most cited h-related index (WoS: 582, Scopus: 705, GS: 1401, MA: 1107), which is also frequently used in evaluation practice, is Leo Egghe's g-index.

> A set of papers has a g-index g if g is the highest rank such that the top g papers have, together, at least g^2 citations. This also means that the top $g+1$ papers have less than $(g+1)^2$ papers. [12.139]

The rationale behind the use of the g-index instead of the h-index is to take the actual citation rate of the top-cited papers into account. At the same time, this is the source of its main weakness: its sensitivity to outliers. The g-index can be equivalently defined as the largest number n of highly cited articles for which the average number of citations is at least n [12.140]. This definition can easily be extended from the arithmetic average to other (geometric, harmonic, etc.) averages [12.141], the median or other statistical measures [12.142]. By choosing the averaging function, the sensitivity of the indicator can be fine-tuned as needed.

12.6 Mathematical Models of the *h*-Index

The introduction of the h-index was met with a certain degree of incomprehension because of the obvious difficulties in categorizing the index into the systematics of orthodox statistics. Some would have excommunicated it as inappropriate, others—just as unjustly—glorifying it as something independent of the traditional bibliometric indicators.

Several papers have attempted to find a place for the h-index in the edifice of mathematics. Others tried to relate it to existing, well-known indicators.

12.6.1 Hirsch's Model

In his original paper, Hirsch gives only very simplistic models of the relation of the h-index to the number of publications and citations. His assumptions are admittedly oversimplified (e. g., constant annual rate of publications of an author or constant annual rate of citations to a paper), and obviously serve as only a rough approximation. Nevertheless, the simple relation $h = (N_{c,\text{tot}}/c)^{1/2}$ (where $N_{c,\text{tot}}$ is the total number of citations and c is an empirical constant found by Hirsch to range between 3 and 5) is often referred to in the literature as the one-parameter Hirsch model of the h-index [12.143–146].

12.6.2 The Lotkaian Framework

The concept of Lotkaian informetrics was introduced by *Leo Egghe* [12.147]. The idea is simple. Since many of the informetric distributions are fat-tailed, i. e., their tails are bounded by an inverse power function (rather than an exponential function, as is the case for most of the distributions in the physical sciences), an informetrics model was constructed where all variables were postulated to have an exact power function distribution, $F(x) = x^{-\alpha}$; $f(x) = \alpha x^{-(\alpha+1)}$. In the discrete case, this is equivalent to *Lotka*'s law [12.148]. In this model, many informetric (or, formally equivalently, bibliometric or scientometric) quantities, relations or laws can be described by simple, closed formulae, which might be considered a zeroth approximation of more complete, multiparameter models.

A critical question in such cases is the structural stability of the model: whether small perturbations of the model may lead to indefinite deviations in the results. If the model is not structurally stable, then no conclusions can be extended beyond its strict limits. We are unaware of any attempt to study the structural stability of Lotkaian informetrics.

Egghe's book was published before Hirsch's paper, so it could not contain the Lotkaian interpretation of the h-index. This was remedied by the paper [12.149]. Using the Lotkaian model, the authors concluded that $h = T^{1/\alpha}$, where T is the total number of sources (in the original example of Hirsch, the number of papers of an author). No empirical test of the result was attempted in the paper.

Later attempts [12.146, 150] revealed that the statistical performance of the Lotkaian model, unsurprisingly, fell behind that of its more sophisticated competitors. In the words of *Burrell*:

> This is not to say that Egghe and Rousseau are mathematically incorrect—they do indeed derive the h-index for the mathematical functions that they consider. The practical problem is that these mathematical functions seem not to be appropriate for the actual context of interest as originally presented by Hirsch. [12.145]

12.6.3 Extreme Value Theory

Glänzel [12.151] recognized that the h-index can be related to the so-called Gumbel characteristic extreme values of a probability distribution.

Extreme value theory is of utmost importance whenever not the *typical* but rather the *exceptional* behavior of a statistical variable is considered. Typical examples are prediction of floods or sport records. Extreme value estimators are not (asymptotically) consistent in the statistical sense: as the sample size increases indefinitely, the resulting sequence of estimates does not converge in probability to a finite value. All records will be broken someday. Being closely related to extreme values, this *inconsistency* of the h-index is an inherent property that evoked initial aversion in some.

Let X be a random variable. In our case, X represents the citation rate of a paper. The probability distribution of X is denoted by $p_k = P(X = k)$ for every $k \geq 0$, and the cumulative distribution function is denoted by $F(k) = P(X < k)$. Put $G_k = G(k) := 1 - F(k) = P(X \geq k)$. Gumbel's r-th characteristic extreme value (u_r) is then defined as

$$u_r := G - 1\frac{r}{n} = \max\left\{k : G(k) \geq \frac{r}{n}\right\},$$

where n is a given sample with distribution F. The theoretical h-index, h^*, can consequently be defined as

$$\begin{aligned} h^* &:= \max\{r : u_r \geq r\} \\ &= \max\left\{r : max\left\{k : G(k) \geq \frac{r}{n}\right\} \geq r\right\}. \end{aligned}$$

If there exists such an index r that $u_r = r$, then we obviously have $H := r$ and we can write $H := u_H$.

Let us now consider an important special case, namely, the discrete Paretian distributions with finite expectation. Most distributions used for modeling publication activity and citation processes belong to this category.

We say that the distribution of a random variable X is Paretian if it asymptotically obeys Zipf's law, i.e., if $\lim_{k\to\infty} G_k k^{-\alpha} = \text{constant}$. Asymptotically Pareto-distributed random variables obviously meet this definition, since $p_k = P(X = k) \approx \gamma(\beta + k)^{-(\alpha+1)}$ if $k \gg 1$; $\alpha > 1$, where β and γ are positive constants. In what follows we will deal with this family of distributions. For $k \gg N$ we obtain $p_k = P(X = k) \approx \gamma k^{-(\alpha+1)}$ and $G_k = P(X \geq k) \approx \gamma_1 k^{-\alpha}$, where γ_1 is a positive constant. Hence, we have

$$E(X) = \sum_{k=0}^{\infty} k p_k = \sum_{k=0}^{\infty} G_k < \infty \quad \text{if} \quad \alpha > 1\,.$$

By elementary manipulation of the cumulative distribution function, we obtain the following approximation from the above definition of Gumbel's r-th characteristic extremes.

$$u_r \approx c_1 \left(\frac{n}{r}\right)^{1/\alpha},$$

where c_1 is a positive constant. Applying the Hirsch condition to this approximation results in the property

$$h^* = u_{h^*} \approx c_1 \left(\frac{n}{h^*}\right)^{1/\alpha}, \quad \text{if} \quad n \gg 1\,.$$

Hence, we have

$$h^* \approx c_2 n^{1/(\alpha+1)}\,, \quad \text{if} \quad n \gg 1\,,$$

where $c_2 = c_1^{\alpha/(\alpha+1)}$ is a positive constant. In verbal terms, the h-index is approximately proportional to the $(\alpha+1)$-th root of the number of publications.

The next question is whether and how the h-index of a journal is determined by the parameters of its citation distribution (first of all, of its expected value, the mean citation rate) and the sample size: the number of papers published in the corresponding journal.

In the case of a two-parameter Pareto distribution with parameters β, α, the expected value, x, is

$$x = \frac{\beta}{\alpha - 1}\,,$$

while the constant, c_1, above is $c_1 = \beta$. Hence,

$$h^* = u_{h^*} \approx c_1 \left(\frac{n}{h^*}\right)^{1/\alpha} \approx \beta^{\alpha/(\alpha+1)} n^{1/(\alpha+1)}\,.$$

In the special case of $\alpha = 2$, which corresponds to a Lotka distribution with exponent 3, we have $x = \beta$; hence, we finally obtain

$$h^* = cn^{1/3}x^{2/3}\,, \tag{12.1}$$

where c is a positive real value of order 1.

Although clearly dependent on several precarious conditions, (12.1) proved to be surprisingly general and stable.

It was first tested on the citation distributions of more than 6000 journals over 2 years (2001 and 2002) from the Web of Science database [12.71]. The empirical and calculated values were strongly correlated ($r^2 \approx 0.95$), and the value of the constant was $c \approx 0.75$.

A comparison of 40 countries [12.68] yielded a correlation with $r^2 \approx 0.99$ and $c = 0.93$.

Later, the model was successfully applied for various samples.

It was tested, for examples:

- On the most actively publishing Moroccan scientists (1997–2006) with $r^2 \approx 0.95$ and $c = 0.91$ [12.152].
- On the top 10 journals and institutions (ranked by h-index) from the Web of Science [12.143]. The author compared three models: that of *Hirsch*'s original paper [12.1], the model of *Egghe* and *Rousseau* [12.149], and the Glänzel–Schubert model. The author found that with $c = 0.9$ for journals and $c = 1.0$ for institutions, the Glänzel–Schubert model gives by far the best fit, with Pearson correlation coefficients of 0.995 and 0.885, respectively. The same author later repeated the comparison on a set of universities based on Essential Science Indicators (ESI) data (1999–2009) [12.144]. He found that

  > the Glänzel–Schubert estimation corresponds best to real data. This is confirmed by calculating both Pearson and Spearman correlation coefficients (as the Pearson coefficient measures a linear relation and the Spearman correlation measures whether ranks correspond, we use both coefficients for synthetic measures) between the estimates and the real h-indices.

  The Pearson coefficient was 0.986 and the Spearman coefficient was 0.976. The value of the constant was not reported.
- On Italian analytical chemists with $r^2 = 0.980$ and $c = 0.80$ [12.153].
- On Malaysian universities with $r^2 = 0.98$ and $c = 0.97$ [12.154].

Some authors varied also the α values to find an optimal fit:

- On a selected set of journals and institutes, the size dependence of the h-index was found to be $n^{0.4}$; that corresponds to $\alpha = 1.5$ [12.155, 156].
- On a sample of 50 pharmacology and 50 psychiatry journals (the top 50 journals of an impact factor-ranked list), the optimal choice was $\alpha = 2.2$, when $r^2 = 0.92$ and $c = 0.71$ [12.85].
- On a sample of 134 ecology and 54 forestry journals, parameters $\alpha = 1.77$; $c = 0.70$ and $\alpha = 1.97$; $c = 0.78$ were found, respectively. Instead of linear correlation, the authors used more sophisticated goodness-of-fit tests to conclude:

  > With regard to their fit, we observe that the Glänzel–Schubert model under a Gaussian distribution provides the best fit to the two sets of data [12.146].

Equation (12.1) serves well for non-citation-based h-type indices too.

For the co-author network h-indices of 36 journals in the field of Dentistry & Oral Medicine, the actual h and estimated h^* index values showed a correlation of $r^2 = 0.56$ and $c = 0.60$ [12.100].

Similar to the co-author network, a network of jazz musicians playing in joint sessions was established [12.125]. The estimated collaboration h-indices of the musicians fitted well to the actual values ($r^2 = 0.88$, $c = 0.81$).

The PartnersHIp Ability Index (PHI, φ) also obeys (12.1). The φ-indices of the 34 awardees of the George Hevesy medal (the premier international award of excellence to honor outstanding achievements in radioanalytical and nuclear chemistry) were determined, and tested with the result: $r^2 = 0.85$, $c = 0.96$ [12.119].

A test on a vastly larger sample was also successful. The co-authorship network of over a million Data Bibliography and Library Project (DBLP) authors was analyzed, and actual and estimated φ-indices were correlated, with the result: $r^2 = 0.87$, $c = 0.67$ [12.121].

A h-type measure of product popularity in e-commerce was also defined and modeled by (12.1) [12.157]. A fairly good fit was found ($r^2 = 0.74$, $c = 0.53$). The fit was substantially improved by changing the α from 2 to 0.6 ($r^2 = 0.91$, $c = 0.35$).

Equation (12.1) was also the starting point for several innovative ideas in h-index-related indicators research.

Prathap proposed using the rough estimate of the h-index, $n^{1/3}x^{2/3}$, as an indicator in its own right. He very aptly called the indicator the *mock* h*-index* [12.135]. It is rather useful when a synthetic indicator is needed for characterizing the number of publications and the citation rate simultaneously, but only the sample size and the average citation rate and not the full citation distribution is known. Although *Prathap* soon renamed the indicator the p-index (for performance) [12.158] and fitted it into a thermodynamic framework [12.159], the basic idea remained unchanged, and found its way to real-life analyses, at least in the author's home country [12.160, 161].

Iglesias and *Pecharromán* built an original theoretical framework based on power law distribution and the stretched exponential distribution in order to compare h-indices among different scientific fields [12.162]. Their result, nevertheless, was admittedly consistent with Glänzel's model. Although the paper is highly cited, real-life practical applications are difficult to find [12.163]. With the explicit use of (12.1), the result of *Iglesias* and *Pecharromán* was later generalized to allow for field normalization, change of time frames or any other changes in measurement scales [12.133].

Glänzel further developed his model by:

- Introducing the z-statistics for the analysis of the tail properties of Pareto-type distributions [12.164]
- Finding the formula

$$\left\{h_1^{(\alpha+1)} + h_2^{(\alpha+1)}\right\}^{\frac{1}{(\alpha+1)}}$$

  for the concatenation of two h-indices with the same α but different sample sizes [12.165]
- Finding the relation of the h-index to the method of characteristic scores and scales [12.166].

12.6.4 Fuzzy Integrals

Torra and *Narukawa* [12.167] found that the h-index could be considered a Sugano integral, a type of so-called fuzzy integral.

Let us consider measures and integrals with respect to a finite reference set $X = \{x_1, \ldots, x_N\}$, e.g., we can consider a set of published works.

Definition 12.1
A set function $\mu\colon 2^X \to \mathbb{R}_+$ is a fuzzy measure if it satisfies the following axioms:

i) $\mu(\emptyset) = 0$ (boundary conditions)
ii) $A \subset B$ implies $\mu(A) \le \mu(B)$ (monotonicity).

A measure is additive when $\mu(A \cup B) = \mu(A) + \mu(B)$ for $A \cap B = \emptyset$.

Definition 12.2
Let X be a reference set. Then the additive fuzzy measure $\mu\colon 2^X \to \mathbb{R}_+$ defined as $\mu(A) = |A|$ is called the counting measure.

Definition 12.3
The Sugeno integral of a function $F\colon X \to \mathbb{R}_+$ with respect to a fuzzy measure μ corresponds to $\max_i \min(f(x_{\sigma(i)}), \mu(A_{\sigma(i)}))$, where $A_\sigma(k) = \{x_{\sigma(i)} | j \le k\}$, and when σ is a permutation such that $f(x_{\sigma(i)}) \ge f(x_{\sigma(i+1)})$ for $i \ge 1$.

Definition 12.4
A researcher has an h-index h if h of his papers have received at least h citations and the rest fewer than h citations.

Using X_r and f as above, this can be expressed as: the h-index of researcher r corresponds to $h_r = \max_i \min(f(x_{\sigma(i)}), i)$, where $\{\sigma(1), \ldots, \sigma(N)\}$ is a permutation of $\{1, \ldots, N\}$ such that $f(x_{\sigma(1)}) \ge f(x_{\sigma(2)}) \ge \ldots \ge f(x_{\sigma(N)})$.

Proposition 12.1
The h-index corresponds to the Sugeno integral of f with respect to the counting measure μ.

The authors claim that the interpretations presented "permit one to apply those results obtained in the field of fuzzy integrals and aggregation to analyze current indexes as well as to define new ones," namely:

1. Index that stresses the importance of active researchers.
2. Index that takes into account publisher credibility. To make matters simple, one may define publisher credibility as the impact factor of the journal.
3. Index that takes into account the impact of related papers.

These potentials do not seem to be realized as yet.

M. Bras-Amorósa et al. [12.168] used the Sugano integral to introduce their unique idea of counting the citations to a paper only from authors who had previously never collaborated with the author(s) of the paper in question. It is not the fault of the Sugano integral concept but rather the difficulties in implementing the indicator and the doubts regarding its interpretation (all honestly revealed in the paper) that prevented its wider adoption.

Mesiar and *Gagolewski* [12.169] recently used the Sugano integral framework to introduce some modifications into the h-index to compensate for its insensitivity to a large number of papers with relatively small numbers of citations and with respect to a large number of citations received by a small number of papers, as well as its complete lack of taking uncited publications into account. They also called attention to the connection of the h-index to another fuzzy measure: "the h-index is the Ky Fan metric computed with respect to the 0 vector".

In summary, fuzzy integrals seem to serve as a rather elegant and general framework within which to find a place for the h-index and related constructions in the edifice of mathematics, but so far without much practical significance.

12.6.5 Axiomatics

It is always a great challenge for mathematicians to find the bare skeletons of an intellectual construction in the form of basic postulates or axioms. Axioms of ranking in general, and citation-based ranking in particular, were well known before the appearance of the h-index [12.170, 171]. The roots of all these attempts lead back to *Arrow*'s work [12.172, 173].

The axiomatization of the h-index apparently began with *Woeginger* [12.174], who formulated "five fairly natural axioms that capture certain desired elementary properties of a scientific impact index $\mathrm{f}: \mathrm{X} \to \mathbb{N}$":

A1. If the $(n+1)$-dimensional vector y results from the n-dimensional vector x by adding a new article with $f(x)$ citations, then $f(y) \leq f(x)$.

A2. If the $(n+1)$-dimensional vector y results from the n-dimensional vector x by adding a new article with $f(x)+1$ citations, then $f(y) > f(x)$.

B. If the n-dimensional vector y results from the n-dimensional vector x by increasing the number of citations of a single article, then $f(y) \leq f(x)+1$.

C. If the n-dimensional vector y results from the n-dimensional vector x by increasing the number of citations of every article by at most one, then $f(y) \leq f(x)+1$.

D. If the $(n+1)$-dimensional vector y results from the n-dimensional vector x by first adding an article with $f(x)$ citations and afterwards increasing the number of citations of every article by at least one, then $f(y) > f(x)$.

In the spirit of Arrow's impossibility theorem, it was proved that "no scientific impact index can simultaneously satisfy all five axioms A1, A2, B, C, and D".

As to the h-index, the main result was the characterization theorem: "A scientific impact index $\mathrm{f}: \mathrm{X} \to \mathbb{N}$ satisfies the three axioms A1, B, and D, if and only if it is the h-index".

As a by-product of the axioms, the author defined the dual of the h-index, referred to as w-index. The duality is elucidated by the pair of definitions:

- The h-index is the scientific impact index $h: X \to \mathbb{N}$ that assigns to vector $x = (x_1, \dots, x_n)$ the value $h(x): \max\{k : x_m \leq k \text{ for all } m \geq k\}$.
- The w-index is the scientific impact index $w: X \to \mathbb{N}$ that assigns to vector $x = (x_1, \dots, x_n)$ the value $w(x) := \max\{k : xm \leq k - m + 1 \text{ for all } m \geq k\}$.

The dual of the theorem concerning the h-index is: "A scientific impact index $\mathrm{f}: \mathrm{X} \to \mathbb{N}$ satisfies the three axioms A1, B, and C, if and only if it is the w-index".

It is important to note that although five axioms are set, the h-index is characterized by only three of them.

In a subsequent paper [12.175], *Woeginger* replaced axiom B in the characterization theorem with the more appealing symmetry axiom (axiom S). For any publication vector x, the "reflected publication vector" $R(x)$ was defined as $R(x)_i = |(k : x_k \geq i)|$. Then,

S. For any $x \in X$ we have $f(x) = f(R(x))$.

The characterization theorems are now as follows:

- A scientific impact index $f: X \to \mathbb{N}$ satisfies the three axioms A1, S, and D, if and only if it is the h-index.
- A scientific impact index $f: X \to \mathbb{N}$ satisfies the three axioms A2, S, and C, if and only if it is the w-index.

Woeginger's axiomatics were further extended to other h-related indices [12.176, 177].

In a paper submitted almost simultaneously with Woeginger's first paper, and therefore obviously unable to acknowledge it, *Marchant* characterized the ranking according to the h-index using six axioms (i. e., it is a ranking according to the h-index if and only if it satisfies these six axioms) [12.178]. He later compared his results with those of *Woeginger* [12.179].

Using Woeginger's monotonicity axiom (axiom D) and two other axioms, *Quesada* characterized the h-index in the domain of all non-negative real-valued indices [12.180]. He subsequently published several characterizations without the monotonicity axiom [12.181, 182].

Hwang retained two of Woeginger's axioms (A1 and B) and replaced monotonicity by expansion consistency [12.183]. He was apparently unaware of Quesada's characterizations without using the monotonicity axiom.

Miroiu summarized several existing characterizations and introduced some new ones [12.184].

12.6.6 Statistical Reliability

A statistical indicator is virtually useless if its reliability cannot be assessed, at least as a rough approximation. The h-index is often criticized for being an integer, and several *improved* versions were proposed to make it a real number to allow for *greater resolution*. No one, however, can answer the question as to whether an index value of 15.3 differs significantly from 15.7. It is not easy to say, either, that an h-index value of 15 differs significantly from 17, but for the h-index there exists some error estimations and reliability tests.

Table 12.11 Confidence intervals for the h-index ($p = 0.95$, $\alpha = 2$)

h-Index	Lower-bound	Upper-bound
20	18	23
25	22	28
30	27	33
35	32	39
40	37	44
50	46	54
100	94	106

Based on the results of *Beirlant* and *Einmahl* [12.185], *Barcza* and *Telcs* [12.186] and his earlier papers [12.164, 166], *Glänzel* determined the confidence interval for an empirical h-index for any confidence level and any α [12.187]. In Table 12.11, the confidence intervals are given for the most generally used confidence level, 0.95, and the typical α value of 2.

It can be seen that the statistical uncertainty of the empirical h-value is decreasing with increasing h-index but not below $\pm 6\%$. The claim to create *higher-resolution* indices is, thus, illusory.

From a similar premise, using a somewhat different statistical apparatus, *Cerchiello* and *Giudici* [12.188] reach a similar conclusion: in the range of h between 10 and 20, at a confidence level 0.90, the confidence interval is about $\pm 25\%$.

Fiorenzo et al. [12.189] used the Egghe–Rousseau model [12.149] and Monte Carlo simulation to find the effect of the error in citation data to the error of the h-index. They found that 1% error in the citation distribution data may cause 3–5% error in the h-index.

12.7 Closing Remarks

In the Web Ontology Language (OWL) [12.190], the *citation function* class contains the following subclasses and members [12.191]:

- Factual function: cites as authority, cites as data source, cites as evidence, cites as metadata document, cites as source document, cites for information, compiles, contains assertion from, documents, gives background to, gives support to, includes excerpt from, includes quotation from, is cited as authority by, is cited as data source by, is cited as evidence by, is cited as metadata document by, is cited as source document by, is cited for information by, is compiled by, is documented by, is plagiarized by, is updated by, obtains background from, obtains support from, plagiarizes, provides assertion for, provides conclusions for, provides data for, provides excerpt for, provides method for, provides quotation for, updates, uses conclusions from, uses data from, uses method in.
- Negative rhetorical function: corrects, critiques, derides, disagrees with, disputes, is corrected by, is critiqued by, is derided by, is disagreed with by, is disputed by, is parodied by, is qualified by, is refuted by, is retracted by, is ridiculed by, parodies, qualifies, refutes, retracts, ridicules.
- Neutral rhetorical function: cites, cites as recommended reading, cites as related, describes, discusses, extends, has reply from, is cited as recommended reading, is cited as related by, is cited by, is

described by, is discussed by, is extended by, is reviewed by, is speculated on by, replies to, reviews, speculates on.
- Positive rhetorical function: agrees with, cites as potential solution, confirms, credits, is agreed with by, is cited as potential solution, is confirmed by, is credited by, is supported by, supports.
- Rhetorical function: cites as authority, cites as evidence, contains assertion from, gives background to, gives support to, is cited as authority by, is cited as evidence by, obtains background from, obtains support from, provides assertion for.

After spending many, many hours studying the h-related literature, the authors of this chapter came to the conclusion that practically all members of this list are represented among the citations to Hirsch's paper [12.1].

Blessing or curse, fad or fashion, Hirsch and his h-index left an indelible mark not only in scientometrics/bibliometrics/informetrics/webometrics, but also in quite distant areas such as network analysis, proteomics or environmental monitoring. The paper and the index thus appear to have a real and lasting impact, and the h-related literature will still provide work for analysts in years and decades to come.

Acknowledgments. One of the authors, András Schubert was supported by the European Commission through FP7 Grant No. 613202 (IMPACT-EV project).

12.A Appendix

Table 12.12 Variants of Hirsch's original paper as cited in the Web of Science (accessed in February, 2017)

Cited author	Cited work	Year	Volume	Issue	Page
Harzing, A. W....Hirsch, J. E.	P NATL ACAD SCI USA	2005	102		16 569
Hirsch, J.	P NATL ACAD SCI USA	2005	102	46	16 569
Hirsch, J. E.	P NATL ACAD SCI USA	2005	102		16 569
HIRSCH J	ARXIVPHYSICS050825	2005			
Hirsch, J.	P NATL ACAD SCI US	2005	102		
HIRSCH J	P NATL ACAD SCI USA	2006	102		16 569
Hirsch, JE	P NATL ACAD SCI USA	2005	102		16 569
Hirsch, J.	P NATL ACAD SCI USA	2005	102		15 572
HIRSCH J	P NATL ACAD SCI USA	2005	46		16 569
HIRSCH J	P NATL ACAD SCI USA	2005			16 569
Hirsch, J. E.	AN INDEX TO QUANTIFY	2005			
HIRSCH JE	ARXHIV	2005	5		1
HIRSCH JE	ARXIV 0213	2005			
HIRSCH JE	ARXIV ORG E PRINT AR	2005			
Hirsch, J. E.	ARXIV PHYSICS 050802	2005			
HIRSCH JE	ARXIVARXIVPHYSICS050	2005			
HIRSCH JE	ARXIVPHYS0508025	2005			
Hirsch, J. E.	ARXIVPHYSICS0508025	2006	5		
Hirsch, J E	ARXIVPHYSICS0508025	2005	5		
Hirsch, Jorge E.	ARXIVPHYSICS0508025	2005			
HIRSCH JE	ARXIVPHYSICS0508025V	2006			
Hirsch, J.E.	ARXIVPHYSICS0508025V	2005			
Hirsch, J.E.	INDEX QUANTIFY INDIV	2005			16 569
Hirsch, JE	INDEX QUANTIFY INDIV	2005			
Hirsch, J. E.	NATL ACAD SCI US	2005	102		16 569
Hirsch, J. E.	P NAT AC SCI US	2005			
HIRSCH JE	P NAT AC SCI US AM	2005	102		
Hirsch, J. E	P NAT AC SCI US PNAS	2005			
Hirsch, J. E.	P NAT ACAD SCI	2005	102		46
Hirsch, J. E.	P NAT ACAD SCI US	2005		46	
Hirsch, J. E.	P NATL ACAD SCI	2005	102	46	
Hirsch, J.E.	P NATL ACAD SCI	2005			
Hirsch, J. E.	P NATL ACAD SCI US	2005			

Table 12.12 (continued)

Cited author	Cited work	Year	Volume	Issue	Page
Hirsch, JE	P NATL ACAD SCI USA	2005	572	46	16
Hirsch, JE	P NATL ACAD SCI USA	2005	572	46	569
HIRSCH JE	P NATL ACAD SCI USA	2005	104		569
HIRSCH JE	P NATL ACAD SCI USA	2005	104		572
Hirsch, J.E.	P NATL ACAD SCI USA	2005	102	46	1
Hirsch, J. E.	P NATL ACAD SCI USA	2005	102	46	4
Hirsch, JE	P NATL ACAD SCI USA	2005	102		16
Hirsch, J.	P NATL ACAD SCI USA	2005	102	46	165
HIRSCH, Jorge E.	P NATL ACAD SCI USA	2005	102		16 569
HIRSCH JE	P NATL ACAD SCI USA	2005	102		1659
HIRSCH JE	P NATL ACAD SCI USA	2005	102		16 509
Hirsch, JE	P NATL ACAD SCI USA	2005	102		16 572
HIRSCH JE	P NATL ACAD SCI USA	2005	46		165
HIRSCH JE	PHYSICS0508025	2005			
Hirsch, JE	PNAS	2005	102	46	16 563
HIRSCH JE	PNAS	2005	102		
HIRSH JE	P NAT AC SCI NOV 15	2005			
Hirsch, J. E	P NATL ACAD SCI USA	2005	102	46	16 569
Jorge, E...Hirsch, JE	P NAT ACAD SCI	2005	102		16 569
NAVARRETE-CORTES, J....HIRSCH, J.	P NATL ACAD SCI US W	2005		102	16 569

12.B Appendix

Table 12.13 h-related indices included in references [12.3, 21] and in [12.22]

Index	Author	Year
Adaptive pure h-index	Chai JC, Hua P-H et al.	2008
A-Index	Jin BH, Liang LM et al.	2007
Amplitude-index	Valentinuzzi ME, Laciar E, Atrio JL	2007
AR-index	Jin B	2007
b-Index	Brown RJC	2009
ch-Index	Ajiferuke I, Wolfram D	2010
Citation speed s-index	Bornmann L, Daniel HD	2010
Citer h-index	Franceschini F, Maisano D et al.	2010
Complementary h-index	Batista PD, Campiteli MG et al.	2005
Contemporary h-index	Sidiropoulos A, Katsaros D, Manolopoulos Y	2007
Dynamic h-type index	Rousseau R, Ye FY	2008
e-Index	Zhang CT	2009
Environment hj-indices	Dorta-Gonzalez P, Dorta-Gonzalez MI	2010
f-Index	Tol RSJ	2009
f-Index	Katsaros D, Akritidis L, Bozanis P	2009
f-Index	Franceschini F, Maisano D	2010
First-author h-index	Opthof T, Wilde AAM	2009
Fractional g-index	Egghe L	2008
Fractional h-index	Egghe L	2008
Freshness indicator	Wu J, Lozano S, Helbing D	2011
Generalized h-index	Glänzel W, Schubert A	2009
g-Index	Egghe L	2006
$h^{(2)}$-Index	Kosmulski M	2006
h'-Index	Hirsch JE	2005
h^2-Lower, h^2-center and h^2-upper	Bornmann L, Mutz R, Daniel HD	2010
Harmonic p-index	Prathap G	2011

Table 12.13 (continued)

Index	Author	Year
h-Bar index	Hirsch JE	2010
h-Core citations	Rousseau R	2006
hg-Index	Alonso S, Cabrerizo FJ et al.	2010
h_{int}-Index	Kosmulski M	2010
h_{m}-Index	Molinari A, Molinari JF	2008
h_{m}-Index	Schreiber M	2008
hpd Index	Kosmulski M	2009
hpy Index	Kosmulski M	2009
hw-Index	Egghe L, Rousseau R	2008
if^2-Index	Boell SK, Wilson CS	2010
Impact vitality indicator	Rons N, Amez L	2009
Iterated weighted c-index	Todeschini R, Baccini A	2016
Iterated weighted g-index	Todeschini R, Baccini A	2016
Iterated weighted h-index	Todeschini R, Baccini A	2016
j-Index	Todeschini R	2011
k-Index	Anania G, Caruso A	2013
k-Index	Ye FY, Rousseau R	2010
Maxprod	Kosmulski M	2007
Mean h-index	Lazaridis T	2010
m-Index	Bornmann L, Mutz R, Daniel HD	2008
Mock h_{m}-index	Prathap G	2010
Modified h-index	Schreiber M	2008
Modified r-index	Panaretos J, Malesios C	2009
m-Quotient	Hirsch JE	2005
${}^{n}h_3$-Index	Vieira ES, Gomes JANF	2010
n-Index	Namazi MR, Fallahzadeh MK	2010
Normalized h-index	Levitt JM, Thelwall M	2007
Normalized h-index	Sidiropoulos A, Katsaros D, Manolopoulos Y	2007
Paper fractional g-index	Egghe L	2008
Paper fractional h-index	Egghe L	2008
p-Index	Prathap G	2010
Pure h-index	Wan JK, Hua PH, Rousseau R	2007
Pure r-index	Wan JK, Hua PH, Rousseau R	2007
q^2-Index	Cabrerizo FJ, Alonso S et al.	2009
Rational g-index	Guns R, Rousseau R	2009
Rational h-index	Ruane F, Tol RSJ	2008
RC- and CC-indices	Abbasi A, Altmann J, Hwang J	2010
R-index	Jin BH, Liang LM, Rousseau R, Egghe L	2007
Role based h-maj-index	Hu XJ, Rousseau R, Chen J	2010
Second-generation citations h-index	Kosmulski M	2010
Selectivity-index	Valentinuzzi ME, Laciar E, Atrio JL	2007
Self-citation-corrected $h^{(2)}$-index	Kosmulski M	2006
Self-citation-corrected h-index	Kosmulski M	2006
Seniority-independent Hirsch-type index	Kosmulski M	2009
Sharpened h-index	Schreiber M	2007
Sharpened modified h-index	Schreiber M	2009
Single paper h-index	Schubert A	2009
Specific-impact s-index	De Visscher A	2010
Tapered h-index	Anderson TR, Hankin KSH, Killworth PD	2008
Time scaled h-index	Mannella R, Rossi P	2013
t-Index	Tol RSJ	2009
Trend h-index	Sidiropoulos A, Katsaros D, Manolopoulos Y	2007

Table 12.13 (continued)

Index	Author	Year
v-Index	Riikonen P, Vihinen M	2008
w-Index	Anania G, Caruso A	2013
w-Index	Wohlin C	2009
w-Index	Wu Q	2010
wl-Index	Wan X, Liu F	2014
x-Index	Wan X	2014
π-Index	Vinkler P	2009
πv-Index	Vinkler P	2010
α-Index	Abt HA	2012

References

12.1 J.E. Hirsch: An index to quantify an individual's scientific research output, Proc. Natl. Acad. Sci. USA **102**(46), 16569–16572 (2005)

12.2 N.K. Vitanov: *Science Dynamics and Research Production. Indicators, Indexes, Statistical Laws and Mathematical Models* (Springer, Cham 2016)

12.3 R. Todeschini, A. Baccini: *Handbook of Bibliometric Indicators. Quantitative Tools for Studying and Evaluating Research* (Wiley-VCH, Weinheim 2016)

12.4 Wikipedia: h-Index, https://en.wikipedia.org/wiki/H-index (last accessed August 2018)

12.5 A.W.F. Edwards: System to rank scientists was pedalled by Jeffreys, Nature **437**, 951 (2005)

12.6 Leviathan: What is your Eddington number?, https://road.cc/content/forum/155509-what-your-eddington-number (2015)

12.7 J. Swindells: Calculate your Eddington number!, https://swinny.net/Cycling/-4687-Calculate-your-Eddington-Number (2016)

12.8 Wikipedia: Eddington number, https://en.wikipedia.org/wiki/Eddington_number (last accessed August 2018)

12.9 J.E. Hirsch, G. Buela-Casal: The meaning of the h-index, Int. J. Clin. Health Psychol. **14**, 161–164 (2014)

12.10 P.T. Frangopol: Indexul Hirsch – Un nou indicator scientometric pentru evaluarea rezultatelor unui cercetător ştiinific (The Hirsch index: A new scientometric indicator for the evaluation of the results of a scientific researcher), Rev. Chim. **56**(12), 1279–1281 (2005)

12.11 T. Braun, W. Glänzel, A. Schubert: A Hirsch-type index for journals, Scientist **19**(22), 8 (2005)

12.12 Wikipedia: Polycyclic group, https://en.wikipedia.org/wiki/Polycyclic_group#Hirsch_length (last accessed August 2018)

12.13 Wikipedia: H-factor, https://en.wikipedia.org/wiki/H-factor (last accessed August 2018)

12.14 M. Beals, L. Gross, S. Harrell: Diversity indices: Shannon's H and E, http://www.tiem.utk.edu/~gross/bioed/bealsmodules/shannonDI.html (2000)

12.15 Wikipedia: Herfindahl index, https://en.wikipedia.org/wiki/Herfindahl_index (last accessed August 2018)

12.16 International Service of Geomagnetic Indices: *ASY/SYM* Indices, http://isgi.unistra.fr/indices_asy.php (2013)

12.17 E.J. Lee, M. Kim, H.S. Kim, M.J. Park, Y.K. Lee, H.J. Kang: Development of a novel quality improvement indicator based on the Hemolysis Index, Ann. Lab. Med. **36**(6), 599–602 (2016)

12.18 K. Shoji, E. Regenbogen, D.Y. Jong, S.M. Blaugrund: H-Index: A new measure of glottal efficiency for the pathologic voice, Laryngoscope **102**(10), 1113–1117 (1992)

12.19 J.E. Hirsch: An index to quantify and individual's scientific research output, https://arxiv.org/abs/physics/0508025 (2005)

12.20 P. Ball: Index aims for fair ranking of scientists, Nature **436**, 900 (2005)

12.21 S. Alonso, F.J. Cabrerizo, E. Herrera-Viedma, F. Herrera: h-Index: A review focused in its variants, computation and standardization for different scientific fields, J. Informetr. **3**(4), 273–289 (2009)

12.22 S. Alonso, F.J. Cabrerizo, E. Herrera-Viedma, F. Herrera, h-Index and Variants, http://sci2s.ugr.es/hindex

12.23 M. Norris, C. Oppenheim: The h-index: a broad review of a new bibliometric indicator, J. Doc. **66**(5), 681–705 (2010)

12.24 L. Zhang, B. Thijs, W. Glänzel: The diffusion of H-related literature, J. Informetr. **5**(4), 583–593 (2011)

12.25 L. Zhang, W. Glänzel, F.Y. Ye: The dynamic evolution of core documents: an experimental study based on h-related literature (2005–2013), Scientometrics **106**(1), 369–381 (2016)

12.26 W. Glänzel, H.F. Moed: Journal impact measures in bibliometric research, Scientometrics **53**(2), 171–193 (2002)

12.27 A. Schubert: Az impakt faktor és akiknek nem kell (The impact factor and those who dislike it), Orv. Hetil. **156**(26), 1065–1069 (2015)

12.28 F. Franceschini, D.A. Maisano: Analysis of the Hirsch index's operational properties, Eur. J. Oper. Res. **203**(2), 494–504 (2010)

12.29 R. Rousseau, C. García-Zorita, E. Sanz-Casado: The h-bubble, J. Informetr. **7**(3), 294–300 (2013)

12.30 E. Garfield: Citation analysis as a tool in journal evaluation, Science **178**(4060), 471–479 (1972)
12.31 P. Jacsó: Google Scholar: The pros and the cons, Online Inf. Rev. **29**(2), 208–214 (2005)
12.32 E. Delgado López-Cózar, N. Robinson-García, D. Torres-Salinas: The Google Scholar experiment: How to index false papers and manipulate bibliometric indicators, J. Assoc. Inf. Sci. Technol. **65**(3), 446–454 (2014)
12.33 P. Jacsó: Google Scholar metrics for publications: The software and content features of a new open access bibliometric service, Online Inf. Rev. **36**(4), 604–619 (2012)
12.34 A.-W. Harzing: Publish or Perish, http://www.harzing.com/pop.htm (2007)
12.35 S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang: Complex networks: Structure and dynamics, Phys. Rep. **424**, 175–308 (2006)
12.36 A.-W. Harzing: Microsoft Academic (search): A phoenix arisen from the ashes?, Scientometrics **108**(3), 1637–1647 (2016)
12.37 A.-W. Harzing, S. Alakangas: Microsoft Academic: Is the Phoenix getting wings?, Scientometrics **110**(1), 371–383 (2017)
12.38 P. Azoulay, J.S. Graff Zivin, G. Manso: Incentives and creativity: Evidence from the academic life sciences, RAND J. Econ. **42**(3), 527–554 (2011)
12.39 A. Serenko, N. Bontis: Global ranking of knowledge management and intellectual capital academic journals, J. Knowl. Manag. **13**(1), 4–15 (2009)
12.40 A. Andrés: *Measuring Academic Research: How to Undertake a Bibliometric Study* (Chandos, Oxford 2009)
12.41 R. Sooryamoorthy: *Transforming Science in South Africa: Development, Collaboration and Productivity* (Palgrave Macmillan, New York 2015)
12.42 P. Vinkler: *The Evaluation of Research by Scientometric Indicators* (Chandos, Oxford 2010)
12.43 InCites: Essential science indicators, http://ipscience-help.thomsonreuters.com/inCites2Live/filterValuesGroup/researchAreaSchema/esiDetail.html (2016)
12.44 Elsevier: Scopus, https://www.elsevier.com/solutions/scopus
12.45 P.F. Svider, Z.A. Choudhry, O.J. Choudhry, S. Baredes, J.K. Liu, J.A. Eloy: The use of the *h*-index in academic otolaryngology, Laryngoscope **123**(1), 103–106 (2013)
12.46 S.H. Zyoud, S.W. Al-Jabi, W.M. Sweileh: Bibliometric analysis of scientific publications on waterpipe (narghile, shisha, hookah) tobacco smoking during the period, Tob. Induc. Dis. **12**(7), 2003–2012 (2014)
12.47 R. Rousseau: New developments related to the Hirsch index, Sci. Focus **1**(4), 23–25 (2006), in Chinese; English version available at http://eprints.rclis.org/7616/
12.48 Q.L. Burrell: On the h-index, the size of the Hirsch core and Jin's *A*-index, J. Informetr. **1**(2), 170–177 (2007)
12.49 W. Glänzel: The role of core documents in bibliometric network analysis and their relation with *h*-type indices, Scientometrics **93**(1), 113–123 (2012)
12.50 M. Stalzer, C. Mentzel: A preliminary review of influential works in data-driven discovery, SpringerPlus **5**, 1266 (2016)
12.51 R. Frodeman: Interdisciplinary research and academic sustainability: managing knowledge in an age of accountability, Environ. Conserv. **38**(2), 105–112 (2011)
12.52 M. Kosmulski: A new Hirsch-type index saves time and works equally well as the original *h*-index, ISSI Newsletter **2**(3), 4–6 (2006)
12.53 S.K. Boell, C.S. Wilson: Journal Impact Factors for evaluating scientific performance: use of *h*-like indicators, Scientometrics **82**(3), 613–626 (2010)
12.54 M. Schreiber: To share the fame in a fair way, h_m modifies h for multi-authored manuscripts, New J. Phys. **10**, 040201 (2008)
12.55 V.R. Meyer: The *h* index – help or hype?, Chimia **63**(1/2), 66–68 (2009)
12.56 Chemistry World: *H*-index ranking of living chemists, http://www.rsc.org/images/H-index%20ranking%20of%20living%20chemists(December%202011)_tcm18-211414.pdf (12 December 2011)
12.57 R. Van Noorden: Hirsch index ranks top chemists, Chemistry World, News (23 April 2007), http://www.rsc.org/chemistryworld/News/2007/April/23040701.asp
12.58 P. Broadwith: End of the road for *h*-index rankings, Chemistry World, News (27 November 2012)
12.59 Ranking Web of Universities: 2610 highly cited researchers ($h > 100$) according to their Google Scholar Citations public profiles, 8th edn., http://www.webometrics.info/en/node/58 (2017)
12.60 A.F.J. Van Raan: Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups, Scientometrics **67**(3), 491–502 (2006)
12.61 M.P. Perme, S.J. Stare, R. Zaucer, M. Zaucer: Comparison of the citation distribution and *h*-index between groups of different sizes, J. Informetr. **6**(4), 712–720 (2012)
12.62 T.C.E. Engels, P. Goos, N. Dexters, E.H.J. Spruyt: Group size, *h*-index, and efficiency in publishing in top journals explain expert panel assessments of research group quality and productivity, Res. Eval. **22**(4), 224–236 (2013)
12.63 R. Arencibia-Jorge, I. Barrios-Almaguer, S. Fernandez-Hernandez, R. Carvajal-Espino: Applying successive H indices in the institutional evaluation: A case study, J. Am. Soc. Inf. Sci. Technol. **59**(1), 155–157 (2008)
12.64 M.P. da Luz, C. Marques-Portella, M. Mendlowicz, S. Gleiser, E.S.F. Coutinho, I. Figueira: Institutional *h*-index: The performance of a new metric in the evaluation of Brazilian Psychiatric Post-graduation Programs, Scientometrics **77**(2), 361–368 (2008)

12.65 T. Lazaridis: Ranking university departments using the mean h-index, Scientometrics **82**(2), 211–216 (2010)
12.66 M.H. Huang: Exploring the h-index at the institutional level: A practical application in world university rankings, Online Inf. Rev. **36**(4), 534–547 (2012)
12.67 C.H. Kuan, M.H. Huang, D.Z. Chen: Cross-field evaluation of publications of research institutes using their contributions to the fields' MVPs determined by h-index, J. Informetr. **7**(2), 455–468 (2013)
12.68 E. Csajbók, A. Berhidi, L. Vasas, A. Schubert: Hirsch-index for countries based on Essential Science Indicators data, Scientometrics **73**(1), 91–117 (2007)
12.69 P. Jacsó: The h-index for countries in Web of Science and Scopus, Online Inf. Rev. **33**(4), 831–837 (2009)
12.70 T. Braun, W. Glänzel, A. Schubert: A Hirsch-type index for journals, Scientometrics **69**(1), 169–173 (2006)
12.71 A. Schubert, W. Glänzel: A systematic analysis of Hirsch-type indices for journals, J. Informetr. **1**(2), 179–184 (2007)
12.72 P.S. Pagel, J.A. Hudetz: Bibliometric analysis of anaesthesia journal editorial board members: correlation between journal impact factor and the median h-index of its board members, Br. J. Anaesth. **107**(3), 357–361 (2011)
12.73 E. Garcia-Pachon, I. Padilla-Navas: Impact factor and h index of Spanish biomedical journals, Med. Clin. **142**(5), 226–227 (2014)
12.74 G. Saad: Exploring the h-index at the author and journal levels using bibliometric data of productive consumer scholars and business-related journals respectively, Scientometrics **69**(1), 117–120 (2006)
12.75 J. Mingers, F. Macri, D. Petrovici: Using the h-index to measure the quality of journals in the field of business and management, Inf. Process. Manag. **48**(2), 234–241 (2012)
12.76 G. Prathap: Correlation between h-index, Eigenfactor™ and Article Influence™ of chemical engineering journals, Curr. Sci. **100**(9), 1276–1276 (2011)
12.77 C.-Y. Yin: Do impact factor, h-index and Eigenfactor™ of chemical engineering journals correlate well with each other and indicate the journals' influence and prestige?, Curr. Sci. **100**(5), 648–653 (2011)
12.78 J.D. Olden: How do ecological journals stack-up? Ranking of scientific quality according to the h index, Ecoscience **14**(3), 370–376 (2007)
12.79 A.-W. Harzing, R. van der Wal: A Google Scholar h-Index for journals: An alternative metric to measure journal impact in Economics and Business, J. Am. Soc. Inf. Sci. Technol. **60**(1), 41–46 (2009)
12.80 J.K. Vanclay: Ranking forestry journals using the h-index, J. Informetr. **2**(4), 326–334 (2008)
12.81 Y. Liu, I.K.R. Rao, R. Rousseau: Empirical series of journal h-indices: The JCR category Horticulture as a case study, Scientometrics **80**(1), 59–74 (2009)
12.82 M. Romero-Torres, L.A. Acosta-Moreno, M.-A. Tejada-Gomez: Use of the h index to rank scientific Latin American journals: Colombia a case study, Rev. Esp. Doc. Cient. **36**(1), 003 (2013)
12.83 A. Doja, K. Eady, T. Horsley, M.D. Bould, J.C. Victor, M. Sampson: The h-index in medical education: An analysis of medical education journal editorial boards, BMC Med. Educ. **14**, 251 (2014)
12.84 P. Bador, T. Lafouge: Comparative analysis of impact factor and h-index for pharmacology journals, Therapie **65**(2), 129–137 (2010)
12.85 P. Bador, T. Lafouge: Comparative analysis between impact factor and h-index for pharmacology and psychiatry journals, Scientometrics **84**(1), 65–79 (2010)
12.86 P. Bador, T. Lafouge: Comparative analysis between impact factor and h-index for psychiatry journals, Can. J. Inf. Libr. Sci. **35**(2), 109–121 (2011)
12.87 G.E. Hunt, M. Cleary, G. Walter: Psychiatry and the Hirsch h-index: The relationship between journal impact factors and accrued citations, Harv. Rev. Psychiatry **18**(4), 207–219 (2010)
12.88 A. Matias Gamez: Assessment of psychology journals: Correlation of impact factor and h-index vs. Latindex criteria, Investig. Bibliotec. **27**(61), 15–27 (2013)
12.89 W. Han, Q. Yu, Y. Wang: Comparative analysis between impact factor and h-index for reproduction biology journals, J. Anim. Vet. Adv. **9**(11), 1552–1555 (2010)
12.90 D.R. Hodge, J.R. Lacasse: Evaluating journal quality: Is the h-index a better measure than impact factors?, Res. Soc. Work Pract. **21**(2), 222–230 (2011)
12.91 B. Minasny, A.E. Hartemink, A. McBratney, H.-J. Jang: Citations and the h index of soil researchers and journals in the Web of Science, Scopus, and Google Scholar, PEERJ **1**, 183 (2013)
12.92 S.B. Bird, K. Kent, R. Church, C. Rosenbaum: Impact factors, h indices, and citation analyses in toxicology journals, Clin. Toxicol. **46**(7), 629–629 (2008)
12.93 J. Hönekopp, J. Kleber: Sometimes the impact factor outshines the H index, Retrovirology **5**, 88 (2008)
12.94 M.G. Banks: An extension of the Hirsch index: Indexing scientific topics and compounds, Scientometrics **69**(1), 161–168 (2006)
12.95 K.M. McIntyre, I. Hawkes, A. Waret-Szkuta, S. Morand, M. Baylis: The H-index as a quantitative indicator of the relative impact of human diseases, PLoS ONE **6**(5), e19558 (2011)
12.96 K.M. McIntyre, C. Setzkorn, P.J. Hepworth, S. Morand, A.P. Morse, M. Baylis: A quantitative prioritisation of human and domestic animal pathogens in Europe, PLoS ONE **9**(8), e103529 (2014)
12.97 R. Cox, K.M. McIntyre, J. Sanchez, C. Setzkorn, M. Baylis, C.W. Revie: Comparison of the h-index scores among pathogens identified as emerging

hazards in North America, Transbound. Emerg. Dis. **63**, 79–91 (2016)
12.98 A.G. Murray, M. Wardeh, K.M. McIntyre: Using the *H*-index to assess disease priorities for salmon aquaculture, Prev. Vet. Med. **126**, 199–207 (2016)
12.99 I. Díaz, M. Cortey, À. Olvera, J. Segalés: Use of *H*-index and other bibliometric indicators to evaluate research productivity outcome on swine diseases, PLoS ONE **11**(3), e0149690 (2016)
12.100 A. Schubert, A. Korn, A. Telcs: Hirsch-type indices for characterizing networks, Scientometrics **78**(2), 375–382 (2009)
12.101 D. Eppstein, E.S. Spiro: The *h*-index of a graph and its application to dynamic subgraph statistics, Lect. Notes Comput. Sci. **5664**, 278–289 (2009)
12.102 A. Korn, A. Schubert, A. Telcs: Lobby index in networks, Physica A **388**(11), 2221–2226 (2009)
12.103 D. Eppstein, E.S. Spiro: The *h*-index of a graph and its application to dynamic subgraph statistics, J. Graph Algorithms Appl. **16**(2), 543–567 (2012)
12.104 D. Eppstein, M.T. Goodrich, D. Strash, L. Trott: Extended dynamic subgraph statistics using *h*-index parameterized data structures, Lect. Notes Comput. Sci. **6508**, 128–141 (2010)
12.105 D. Eppstein, M.T. Goodrich, D. Strash, L. Trott: Extended dynamic subgraph statistics using *h*-index parameterized data structures, Theor. Comput. Sci. **447**, 44–52 (2012)
12.106 S. Hartung, C. Komusiewicz, A. Nichterlein, O. Suchý: On structural parameterizations for the 2-club problem, Discrete Appl. Math. **185**, 79–92 (2015)
12.107 C. Komusiewicz, M. Sorge: An algorithmic framework for fixed-cardinality optimization in sparse graphs applied to dense subgraph problems, Discrete Appl. Math. **193**, 145–161 (2015)
12.108 S.K.A. Khan, R.J. Mondragón, L.N. Tokarchuk: Lobby influence: Opportunistic forwarding algorithm based on human social relationship patterns. In: *Proc. 2012 IEEE Int. Conf. Pervasive Comput. Commun. Workshops* (2012) pp. 211–216
12.109 J. Wang, W. Peng, F.-X. Wu: Computational approaches to predicting essential proteins: A survey, Proteomics Clin. Appl. **7**(1/2), 181–192 (2013)
12.110 A. Telcs, M. Csernai, A. Gulyás: Load balanced diffusive capture process on homophilic scale-free networks, Physica A **392**(3), 510–519 (2013)
12.111 I. Kilanioti: Improving multimedia content delivery via augmentation with social information: The social prefetcher approach, IEEE Trans. Multimed. **17**(9), 1460–1470 (2015)
12.112 R. Rousseau, S.X. Zhao: A general conceptual framework for characterizing the ego in a network, J. Informetr. **9**(1), 145–149 (2015)
12.113 L. Lü, T. Zhou, Q.-M. Zhang, H.E. Stanley: The *H*-index of a network node and its relation to degree and coreness, Nat. Commun. **7**, 10168 (1916)
12.114 S.X. Zhao, R. Rousseau, F.Y. Ye: *h*-Degree as a basic measure in weighted networks, J. Informetr. **5**(4), 668–677 (2011)
12.115 S.X. Zhao, P.L. Zhang, J. Li, A.M. Tan, F.Y. Ye: Abstracting the core subnet of weighted networks based on link strengths, J. Assoc. Inf. Sci. Technol. **65**(5), 984–994 (2014)
12.116 S.S. Li, R. Rousseau, F.Y. Ye: Motifs in weighted networks and their Hirsch subgraphs, Malays. J. Libr. Inf. Sci. **21**(3), 21–34 (2016)
12.117 A. Abbasi: *h*-Type hybrid centrality measures for weighted networks, Scientometrics **96**(2), 633–640 (2013)
12.118 S.X. Zhao, F.Y. Ye: Exploring the directed *h*-degree in directed weighted networks, J. Informetr. **6**(4), 619–630 (2012)
12.119 A. Schubert: A Hirsch-type index of co-author partnership ability, Scientometrics **91**(1), 303–308 (2012)
12.120 R. Rousseau: Comments on "A Hirsch-type index of co-author partnership ability", Scientometrics **91**(1), 309–310 (2012)
12.121 G. Cabanac: Experimenting with the partnership ability φ-index on a million computer scientists, Scientometrics **96**(1), 1–9 (2013)
12.122 W. Glänzel: Analysis of co-authorship patterns at the individual level, TransInformação **26**(3), 229–238 (2014)
12.123 D. Besancenot, K.V. Huynh, F. Serranito: Determinant of co-authorship in economics: The French case, Econ. Bull. **35**(1), 680–693 (2015)
12.124 Z. Sabaghinejad, F. Osareh, F. Baji, P.P. Mohammadi: Estimating the partnership ability of Scientometrics journal authors based on WoS from 2001 to 2013 according to φ-index, Scientometrics **109**(1), 73–84 (2016)
12.125 A. Schubert: Jazz discometrics – A network approach, J. Informetr. **6**(4), 480–484 (2012)
12.126 X. Yan, L. Zhai, W. Fan: *C*-index: A weighted network node centrality measure for collaboration competence, J. Informetr. **7**(1), 223–239 (2013)
12.127 L. Zhai, X. Yan, G. Zhang: A centrality measure for communication ability in weighted network, Physica A **392**(23), 6107–6117 (2013)
12.128 V. Gómez, A. Kaltenbrunner, V. López: Statistical analysis of the social network and discussion threads in Slashdot. In: *Proc. 17th Int. Conf. World Wide Web* (2008) pp. 645–654
12.129 D.M. Romero, W. Galuba, S. Asur, B.A. Huberman: Influence and passivity in social media, Lect. Notes Comput. Sci. **6913**, 18–33 (2011)
12.130 D. King, D. Ramirez-Cano, F. Greaves, I. Vlaev, S. Beales, A. Darzi: Twitter and the health reforms in the English National Health Service, Health Policy **110**(2/3), 291–297 (2013)
12.131 P. Cerchiello, P. Giudici: How to measure the quality of financial tweets, Qual. Quant. **50**(4), 1695–1713 (2016)
12.132 A. Ginebreda, A. Jelić, M. Petrović, M. López de Alda, D. Barceló: New indexes for compound prioritization and complexity quantification on environmental monitoring inventories, Environ. Sci. Pollut. Res. **19**(4), 958–970 (2012)
12.133 A. Schubert: Rescaling the *h*-index, Scientometrics **102**(2), 1647–1653 (2015)

12.134 A. Schubert: X-centage: a Hirsch-inspired indicator for distributions of percentage-valued variables and its use for measuring heterodisciplinarity, Scientometrics **102**(1), 307–313 (2015)
12.135 G. Prathap: Is there a place for a mock *h*-index?, Scientometrics **84**(1), 153–165 (2010)
12.136 A. Miroiu: How many athletes qualified for the London Olympic Games?, J. Am. Soc. Inf. Sci. Technol. **65**(8), 1728 (2014)
12.137 L. Waltman: A review of the literature on citation impact indicators, J. Informetr. **10**(2), 365–391 (2016)
12.138 D. Westreich: Editorial: A lasting impact, Am. J. Epidemiol. **181**(11), 829–831 (2015)
12.139 L. Egghe: Theory and practise of the *g*-index, Scientometrics **69**(1), 131–152 (2006)
12.140 Wikipedia: *g*-index, https://en.wikipedia.org/wiki/G-index (last accessed August 2018)
12.141 R.S.J. Tol: The *h*-index and its alternatives: An application to the 100 most prolific economists, Scientometrics **80**(2), 319–326 (2009)
12.142 W. Glänzel, A. Schubert: Hirsch-type characteristics of the tail of distributions. The generalised *h*-index, J. Informetr. **4**(1), 118–123 (2010)
12.143 F.Y. Ye: An investigation on mathematical models of the *h*-index, Scientometrics **81**(2), 493–498 (2009)
12.144 F.Y. Ye: A unification of three models for the *h*-index, J. Am. Soc. Inf. Sci. Technol. **62**(1), 205–207 (2011)
12.145 Q.L. Burrell: Formulae for the *h*-Index: A lack of robustness in Lotkaian informetrics?, J. Am. Soc. Inf. Sci. Technol. **64**(7), 1504–1514 (2013)
12.146 C. Malesios: Some variations on the standard theoretical models for the *h*-index: A comparative analysis, J. Assoc. Inf. Sci. Technol. **66**(11), 2384–2388 (2015)
12.147 L. Egghe: *Power Laws in the Information Production Process: Lotkaian Informetrics* (Elsevier, Oxford 2005)
12.148 A.J. Lotka: The frequency distribution of scientific productivity, J. Wash. Acad. Sci. **16**(12), 317–324 (1926)
12.149 L. Egghe, R. Rousseau: An informetric model for the Hirsch-index, Scientometrics **69**(1), 121–129 (2006)
12.150 C. Malesios: Measuring the robustness of the journal *h*-index with respect to publication and citation values: A Bayesian sensitivity analysis, J. Informetr. **10**(3), 719–731 (2016)
12.151 W. Glänzel: On the *h*-index – A mathematical approach to a new measure of publication activity and citation impact, Scientometrics **67**(2), 315–321 (2006)
12.152 H. Bouabid, B.R. Martin: Evaluation of Moroccan research using a bibliometric-based approach: investigation of the validity of the *h*-index, Scientometrics **78**(2), 203–217 (2009)
12.153 A. Annibaldi, C. Truzzi, S. Illuminati, G. Scarponi: Scientometric analysis of national university research performance in analytical chemistry on the basis of academic publications: Italy as case study, Anal. Bioanal. Chem. **398**(1), 17–26 (2010)
12.154 M. Tahira, R.A. Alias, A. Bakri: Scientometric assessment of engineering in Malaysian universities, Scientometrics **96**(3), 865–879 (2013)
12.155 J.-F. Molinari, A. Molinari: A new methodology for ranking scientific institutions, Scientometrics **75**(1), 163–174 (2008)
12.156 A. Molinari, J.-F. Molinari: Mathematical aspects of a new criterion for ranking scientific institutions based on the *h*-index, Scientometrics **75**(2), 339–356 (2008)
12.157 B. Shao, F. Xiao, S.X. Zhao: Exploring the *h*-type measure and its theoretical model in the context of e-commerce, Curr. Sci. **110**(12), 2276–2280 (2016)
12.158 G. Prathap: The 100 most prolific economists using the *p*-index, Scientometrics **84**(1), 167–172 (2010)
12.159 G. Prathap: An iCE map approach to evaluate performance and efficiency of scientific production of countries, Scientometrics **85**(1), 185–191 (2010)
12.160 K.P. Raghuraman, R. Chander, G. Madras: Scientometric analysis of some disciplines: Comparison of Indian institutions with other international institutions, Curr. Sci. **99**(5), 577–587 (2010)
12.161 M. Pathak, K.A. Bharati: Botanical survey of India (1971-2010): A scientometric analysis, Curr. Sci. **106**(7), 964–971 (2014)
12.162 J.E. Iglesias, C. Pecharromán: Scaling the *h*-index for different scientific ISI fields, Scientometrics **73**(3), 303–320 (2007)
12.163 L. Czarnecki, M.P. Kaźmierkowski, A. Rogalski: Doing Hirsch proud; shaping *H*-index in engineering sciences, Bull. Pol. Acad. Sci. Tech. Sci. **61**(1), 5–21 (2013)
12.164 W. Glänzel: On some new bibliometric applications of statistics related to the *h*-index, Scientometrics **77**(1), 187–196 (2008)
12.165 W. Glänzel: *H*-index concatenation, Scientometrics **77**(2), 369–372 (2008)
12.166 W. Glänzel: The role of the *h*-index and the characteristic scores and scales in testing the tail properties of scientometric distributions, Scientometrics **83**(3), 697–709 (2010)
12.167 V. Torra, Y. Narukawa: The *h*-index and the number of citations: Two fuzzy integrals, IEEE Trans. Fuzzy Syst. **16**(3), 795–797 (2008)
12.168 M. Bras-Amorósa, J. Domingo-Ferrera, V. Torra: A bibliometric index based on the collaboration distance between cited and citing authors, J. Informetr. **5**(2), 248–264 (2011)
12.169 R. Mesiar, M. Gagolewski: *H*-index and other Sugeno integrals: some defects and their compensation, IEEE Trans. Fuzzy Syst. **24**(6), 1668–1672 (2016)
12.170 W.J. Walker: Ranking functions and axioms for linear orders, Discrete Math. **67**(3), 299–306 (1987)
12.171 I. Palacios-Huerta, O. Volij: The measurement of intellectual influence, Econometrica **72**(3), 963–977 (2004)
12.172 K.J. Arrow: A difficulty in the concept of social welfare, J. Polit. Econ. **58**(4), 328–346 (1950)

12.173 K.J. Arrow: *Social Choice and Individual Values* (Wiley, Hoboken 1951)
12.174 G.J. Woeginger: An axiomatic characterization of the Hirsch-index, Math. Soc. Sci. **56**(2), 224–232 (2008)
12.175 G.J. Woeginger: A symmetry axiom for scientific impact indices, J. Informetr. **2**(4), 298–303 (2008)
12.176 R. Rousseau: Woeginger's axiomatisation of the h-index and its relation to the g-index, the $h^{(2)}$-index and the R^2-index, J. Informetr. **2**(4), 335–340 (2008)
12.177 G.J. Woeginger: An axiomatic analysis of Egghe's g-index, J. Informetr. **2**(4), 364–368 (2008)
12.178 T. Marchant: An axiomatic characterization of the ranking based on the h-index and some other bibliometric rankings of authors, Scientometrics **80**(2), 327–344 (2009)
12.179 D. Bouyssou, T. Marchant: An axiomatic approach to bibliometric rankings and indices, J. Informetr. **8**(3), 449–477 (2014)
12.180 A. Quesada: Monotonicity and the Hirsch index, J. Informetr. **3**(2), 158–160 (2009)
12.181 A. Quesada: More axiomatics for the Hirsch index, Scientometrics **82**(2), 413–418 (2010)
12.182 A. Quesada: Further characterizations of the Hirsch index, Scientometrics **87**(1), 107–114 (2011)
12.183 Y.-A. Hwang: An axiomatization of the Hirsch-index without adopting monotonicity, Appl. Math. Inf. Sci. **7**(4), 1317–1322 (2013)
12.184 A. Miroiu: Axiomatizing the Hirsch index: Quantity and quality disjoined, J. Informetr. **7**(3), 10–15 (2013)
12.185 J. Beirlant, J.H.J. Einmahl: Asymptotics for the Hirsch index, Scand. J. Stat. **37**(3), 355–364 (2010)
12.186 K. Barcza, A. Telcs: Paretian publication patterns imply Paretian Hirsch index, Scientometrics **81**(2), 513–519 (2009)
12.187 W. Glänzel: On reliability and robustness of scientometrics indicators based on stochastic models. An evidence-based opinion paper, J. Informetr. **4**(3), 313–319 (2010)
12.188 P. Cerchiello, P. Giudici: On a statistical h index, Scientometrics **99**(2), 299–312 (2014)
12.189 F. Fiorenzo, M. Domenico, M. Luca: The effect of database dirty data on h-index calculation, Scientometrics **95**(3), 1179–1188 (2013)
12.190 S. Bechhofer, F. van Harmelen, J. Hendler, I. Horrocks, D.L. McGuinness, P.F. Patel-Schneider, L.A. Stein: OWL Web Ontology Language Reference. W3C Recommendation 10 February 2004, https://www.w3.org/TR/owl-ref/#References
12.191 A. Nuzzolese, A. Di Iorio, S. Peroni: Functions of Citations, http://www.essepuntato.it/lode/http://www.essepuntato.it/2013/03/cito-functions (2013)

András Schubert

Library and Information Center
Hungarian Academy of Sciences
Budapest, Hungary
schuba@iif.hu

András Schubert, former Head of the Department of Scientometrics and Science Policy at the Library of the Hungarian Academy of Sciences, graduated as chemical engineer from the Budapest University of Technology and Economics. He switched to scientometrics in 1980. He is editor of the journal *Scientometrics*. Since retirement, he devotes much of his time to playing the clarinet and to literary writing, largely children's stories.

Gábor Schubert

University Library
Stockholm University
Stockholm, Sweden

Gábor Schubert studied chemical engineering at the Budapest University of Technology and Economics before becoming engaged with research in computational chemistry. In cooperation with his father, András Schubert, he has been involved in bibliometric studies since the mid-1990s. He is currently Head Bibliometrician of Stockholm University Library.

13. Citation Classes: A Distribution-based Approach for Evaluative Purposes

Wolfgang Glänzel, Bart Thijs, Koenraad Debackere

In this chapter, we describe a scientometric assessment tool that was first introduced as early as the second half of the 1980s, but due to the high computational requirements at that time, the method fell undeservedly into oblivion. The method is called Characteristic Scores and Scales (CSS) and is aimed at providing a more detailed picture of citation impact, with particular regard to the high end of performance. More than two decades after its introduction, the method experienced a revival as a consequence of the burning need for improved and versatile assessment tools, facilitated by the rapid development of information technology and the broad access to electronic data sources.

The first part of this chapter will describe the model, its background and the statistical properties underlying this approach, while the following sections will deal with its implementation within the framework of research evaluation at different levels of aggregation and in various disciplinary and multidisciplinary contexts. Special attention is paid to the applicability to various aggregation levels, such as national research performance, the comparative analysis of institutional research output, as a tool to assist the assessment of individual researchers and as journal impact measures. A graphical sketch of possible applications is used as a road map throughout the chapter to navigate the various methodological issues and fields of use. The chapter begins with a review of previous work, but also aims at presenting new insights and applications in a systematic manner. In addition to the presentation of new results, future perspectives and possible applications of this model within and outside traditional scientometrics will be sketched and highlighted.

Part B | 13

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_13

13.1 General Introduction: The Need for Multilevel Profiling of Citation Impact

The continuing debate on the need for normalization of bibliometric indicators, and the experienced observation of phenomena that lead to serious biases, even distortions, of citation indicators at the institutional, research group and individual levels, has prompted bibliometricians to reform the traditional basic indicator system. The insufficiency in using a single indicator to depict quantitative aspects of research performance has long been experienced in our field. In particular, the disparity between standard models of regular performance versus outstanding performance as reflected by citation impact [13.1, 2] is another phenomenon that renders the expression of performance by a single indicator conceptually impossible. The objectives of improvement are thus to replace the concept of the *linearly structured* indicator model by more complex performance profiles, while ensuring the seamless integration of measures of outstanding and even extreme performance into the standard tools of scientometric performance assessment [13.3].

Therefore, a reduction of the original citation distribution to performance classes on the basis of *Characteristic Scores and Scales* (CSS), as introduced by *Glänzel* and *Schubert* [13.4], has been proposed as an alternative parameter-free solution for performance assessment. Before we introduce this method and describe its properties and various fields of application, we first have a look at the conceptual background of scientometric indicators, pointing to the main opportunities and limitations, followed by the mathematical basis required for the necessary soundness of the underlying methodology.

13.1.1 The Conceptual Background

When turning to quantitative methods in science studies, i. e., statistics, measures and indicators, to capture and express important and characteristic aspects on the basis of sufficiently large numbers of objects to analyze, we are usually faced with several important methodological and practical issues. Among these issues, we must stress the matter of *conceptual* robustness, reliability and validity. From a conceptual viewpoint, these are different from the statistical interpretation. First of all, these issues are related to the design of bibliometric methods, measures and tools, and imply that indicators can really be assumed to measure what they are intended to measure and are, by definition, not sensitive to marginal changes in the system or in the interpretation of data. *Bookstein* [13.5] has noted three (out of other) demons of measurement that are challenges to quantitative approaches in studies of scientific communication. All three of them have strong conceptual roots—and, of course, also statistical implications, as we will see later. The first demon refers to *randomness* (otherwise, a precondition of mathematical statistics), which concerns not only the variables to be measured but also the conditions of measuring and the *environment* of the variables. Scientific communication is subject to the influence of many intra- and extra-scientific factors that are difficult to capture, depict and model by mathematical methods. *Fuzziness* is reaching beyond probabilistic uncertainty; it is related to the impossibility of accurately representing concepts by traditional variable sets. Many properties that are assumed and investigated by quantitative science studies are only partially valid, that is, the assumption of considering full or absolute validity might bias the studies. The relevance of retrieved documents, the determination of (co-)authorship credits, and the subject assignment of publications might serve just as some typical and commonly acknowledged examples of fuzziness. According to Bookstein, *ambiguity* pushes the idea of fuzziness to its extreme. He argues that the conceptual basis for measurement is often weak, although we aim for accurate measurement and the development of techniques to derive precise results and statements, by often searching for relationships with similarly weak concepts. Finally, the adoption of related methods from other disciplines with similar concepts and changes in the application focus, such as the shift of bibliometrics from scientific information to research evaluation and science policy, increases the impact of ambiguity to the detriment of the required precision. According to Bookstein, all mathematical relationships are approximate. One way to cope with ambiguity is to create and apply models that are less sensitive to imperfect conceptualization. Following his ideas, it may also be important in this context to create and apply specific informetric laws that are based on probabilistic models and to ensure that uncritical adoption of models from other disciplines with similarly imperfect conceptual background is reduced to the possible minimum.

Apart from the above demons, which are also shared with other disciplines—above all in the social sciences and humanities—there are some specific challenges in bibliometrics that are related to the dissemination of information. Unlike interaction of matter as in the natural sciences, or trading of products, information need not be created again in order to transfer again, while dissemination also becomes increasingly independent of specific carriers. This effect, together with other particularities of social processes, results in two further demons to bibliometric measures. The first concerns a certain incommensurability of *masses* and elite as shown, e. g., by *Glänzel* and *Schubert* [13.4] and *Glänzel* [13.2] in the case of citation distributions. This effect has strong implications for finding general regularities and setting reference standards. The second issue relates to *outliers* that sometimes blow all reasonable limits and can be understood in the context of the *anathema* of seemingly unlimited dissemination of information, an effect that is also known in datamining [13.6]. *Waltman* et al. [13.7] have reported such observations in bibliometrics in the case of individual citation rates. As a consequence, otherwise statistically marginal numbers of observations can seriously bias, even distort, the reliability of indicators and thereby affect the rank statistics. Such phenomena, of course, compromise the usefulness of common traditional indicators in the sense of applicability and meaningfulness. At this point, conceptual issues turn into mathematical tasks and require concrete mathematical solutions. The effect of extreme observation, for instance, is not only a bibliometric issue. So-called censored data or extreme observations distorting empirical distributions are known in several other fields, e. g., in insurance mathematics [13.8], just to mention one example. This makes it possible to borrow methods from other, even unrelated, disciplines in a critical way and to search for more informetric solutions, the groundwork of which is to be discussed in the following subsection.

13.1.2 The Mathematical Background

The majority of traditional indicators in bibliometrics are based on simple statistical functions and some arithmetic relations of those. These include, for instance, sample means, shares and percentiles. Yet, the distributions underlying publication activity, co-authorship, citation impact and other bibliometric phenomena to be studied have specific properties that challenge the usefulness of these simple statistics in those contexts that have strong policy implications and financial consequences. In other words, reliability and validity questions emerge if (but not only if) the results of bibliometric studies are applied to the evaluation of research performance at various levels, to the allocation of research funding and the promotion of scientists. And, if bibliometric tools have an effect on decision-making in science policy and research management, this might have measurable repercussions on the scientists' behavior, notably in their scholarly communication, as has been described by *Glänzel* and *Debackere* [13.9].

Coming back to the properties that are partially at the roots of the above-mentioned issues, we must emphasize two specific features of bibliometric distributions. The first important property is that bibliometric distributions are typically (nonnegative) integer values with low medians and means, which prevents us from the approximation by continuous distribution models, since any small deviation from these two statistical functions can almost be considered qualitative rather than infinitesimal. This makes for a clear distinction between bibliometric distributions and those in other disciplines such as meteorology, sports, demographics, insurance mathematics or income distributions. The second property has to do with skewness and extreme values. Informetric distributions are characterized by having heavy (fat) tails and large extreme values, often with true outliers. In such situations, the use of means and medians raises serious issues. The median suffers from low discriminatory power because of the first property, and the mean is, although formally a legitimate statistic in these cases as well [13.10], clearly not representative for the underlying sample, because of the second property. These two specific features are accompanied by a third, not typically informetric issue, namely that of multidimensionality. A single indicator can hardly describe most bibliometric phenomena [13.11]. Citation distributions with practically the same mean (or median) may have different shapes, with different shares of cited papers [13.12]. Therefore, single-number indicators should preferably be replaced by indicator groups in order to obtain a more realistic insight, of course, with severe implications for ranking exercises, because these rely by nature on a linear approach [13.13].

Before we switch to the groundwork of setting up the performance classes for profiling citation impact, we refer to one of the bibliometric demons to measurement mentioned in the context of the *conceptual imperfections*. When dealing with the heavy tails of citation distributions, we often encounter a phenomenon that complicates the estimation of parameters considerably. We have observed significant imparity between standard models developed for describing *regular* performance, i. e., the citation impact attracted by the overwhelming majority of the authors and publications versus the ones designed to capture the outstanding performance achieved by a small elite that apparently obeys its own rules [13.2, 4]. And in this context we do not even speak of outliers. Pragmatically, we would need two different values for the same parameter of the distribution, one for the head and body, and one for the tail. *Glänzel* [13.2] therefore used mathematical-statistical methods such as QQ plots and extreme value distributions to analyze to what extent the high end of citation impact, as reflected by the tail of scientometric distributions, is in line with the *standard* citation impact attracted by the majority of scientific papers, and to what extent extreme values and outliers might be responsible for possible deviations and distortions of citation indicators [13.7].

While in several other fields outliers can be discarded as being exceptions, or can be treated as censored data and re-estimated by more realistic ones, in bibliometrics these extreme observations represent the high end of research performance and deserve special attention. This need argues against alternative solutions such as discarding or correcting extreme values. One possible solution is the use of QQ plots and tail parameters to supplement traditional citation-based performance indicators [13.2, 14]. The analysis of the tail can practically be uncoupled from the overwhelming rest of the empirical distribution. The estimation of the tail parameter from the Pareto model, which is most commonly used in scientometrics, can directly be obtained from subsets of order statistics. This approach also allows the construction of confidence intervals for its estimator. Nevertheless, as the above-mentioned studies have pointed to, the estimation of the tail index remains rather problematic, since most methods are still sensitive to the cut-off point for the tail. This is the reason why mathematicians have sought alternative and more robust solutions. This has, of course, strong implications for alternative methods and indicators that take this disproportion into account. Since already minute changes of the tail parameter might have significant consequences in an evaluative context, the recommendation in the study by *Glänzel* [13.2] was to favor a parameter-free solution for the assessment of outstanding performance. This might also help avoid parameter conflicts resulting from estimating parameters on the basis of different parts of the distributions, notably of the low and high ends of the distribution.

Given the above discourse on the multiple issues in bibliometric measurement that can result in biased, even heavily distorted and "interference-prone" indicators, we briefly summarize the desired properties of an alternative mathematical/informetric method that goes beyond the traditional metrics.

The proposed method should:

1. Provide a parameter-free solution for different performance classes
2. Avoid arbitrary or preset thresholds
3. Not be sensitive to or distorted by ties
4. Replace the concept of *linearly structured* indicators by performance profiles
5. Be applicable to various levels of aggregation and gauge observations against several standards
6. Ensure the seamless integration of measures of outstanding and even extreme performance into the standard tools of scientometric performance assessment.

In order to develop a method with such properties, two basic approaches could provide the required mathematical basis. One possibility is the use of quantiles. A solution using six percentile classes was proposed by *Leydesdorff* et al. [13.15]. Their model assumed a set of six rank percentages, which are derived from a reference distribution based on the entire population. The individual publications of the sample under study are then assigned to the corresponding percentile on the basis of their observed citations rates. This method is apparently insensitive to extreme values and outliers, since those will become just one object in the highest percentile class, that is, extreme values and outliers have the same score as any other highly cited paper independently of their actual citation rates. However, two problems arise from this approach, particularly the arbitrariness of preset percentiles and the ties present in both the reference distribution and the observations. *Waltman* and *Schreiber* [13.16] recently found a solution for solving the latter problem and for avoiding the otherwise inevitable ties.

Another solution is a reduction of the original citation distribution, where the individual observations—or, using the parlance of probability theory, the individual events—are grouped into a given number of classes. As a consequence, the distribution of citations over classes instead of individual values is considered. Similar to the previous approach, the reference distribution of the total population or a relevant subpopulation can be used to define the classes and to set the corresponding standards. If such assignment could be made self-adjusting and widely independent of the influence of the various factors typically affecting citation rates, one would obtain an almost universal tool for the evaluation of citation impact. The structure and properties of such a method will be described in the following section.

13.2 The Method of Characteristic Scores and Scales (CSS)

Grouping events and empirical observations into a limited number of certain classes can be done in different ways. In order to understand the background of our solution, we have to go back to some basics and results in probability and distribution theory. Therefore, we will briefly recall the probability-theory-related rudiments before we discuss their detailed statistical implementation.

13.2.1 The Stochastic Model Behind Characteristic Scores and Scales

Let X be a nonnegative random variable. If furthermore X takes, for instance, integer values, the variable might stand for publication activity or citation rates. Mathematicians have proven that conditional expectations, particularly left-truncated moments and their arithmetical combinations, under certain conditions, characterize the underlying distribution. In this context we refer, above all, to theorems by *Hamdan* [13.17], *Kotz* and *Shanbhag* [13.18] and *Glänzel* et al. [13.19]. According to a fundamental representation theorem for probability distributions, the conditional expectations $E(h(X)|X \geq x)$ with $x \geq 0$, where h is a real function defined on the support of X, characterize a large family of continuous and discrete distributions. Some special cases are of particular interest—for instance, the simplest case, if h is the identity function, i. e., $h(x) = x$, $E(X|X \geq x)$ is a linear function of x, if and only if X has an exponential or Pareto distribution of the second kind (also called Lomax distribution), or, in the discrete case, if, and only if X has a geometric or Waring distribution. Hence, the following idea arose: If this property holds for all nonnegative values x, then the property must also be true for particular x values, such as the expected value $b := E(x)$. Hence, we derived the following iteration

$$
\begin{aligned}
b_1 &:= b = E(X|X \geq 0)\;, \\
b_2 &:= E(X|X \geq b_1)\;, \\
b_3 &:= E(X|X \geq b_2)\;.
\end{aligned}
$$

Putting $b_0 := 0$, we have $b_k := E(X|X \geq b_{k-1})$ for all nonnegative integer values $k = 1, 2, \ldots$ For a Lomax/Waring distribution we obtain a very interesting property, specifically

$$
\begin{aligned}
b_k &= E(X|X \geq b_{k-1}) = ab_{k-1} + b_1 \\
&= N(a-1)\left(\sum_{i=0}^{k-1} a^i\right) = N(a^k - 1) \\
\text{with } a &= \frac{\alpha}{\alpha - 1}\;,
\end{aligned}
\tag{13.1}
$$

where N and α are the parameters of the distribution [13.4, 20] and $b_k = E(X|X \geq b_{k-1}) = b_{k-1} + b_1 = kb$ for the exponential/geometric case which is the limiting case of the previous distributions.

In other words, the iteration of truncating the distribution at its expectation to determine the following value results in a geometric series. This property could, of course, be used for parameter estimation, but because of the above-mentioned tail-parameter inconsistency (disproportion between trunk and tail), we refrain from this option. Instead, we use the b_k values to determine intervals that can be used to reduce the original distribution to a limited "state space" consisting of a finite number of classes. If we nevertheless talk about parameters, we then use those to point to some fundamental, albeit approximate, properties. The following one is maybe the most interesting. In a previous paper, *Glänzel* [13.20] showed that for the Lomax distribution, the class sizes depend only on the tail parameter and, in the case of the exponential distribution, which is a limiting case of the Lomax distribution, class sizes are independent of any parameter. In particular, for Lomax, we have the following property

$$
\begin{aligned}
&G(b_k) = 1 - F(b_k) = a^{-k\alpha} = \left(1 - \frac{1}{\alpha}\right)^{k\alpha} = q^k\;, \\
&\quad k = 1, 2, 3, \ldots \quad \text{with } q := a^{-\alpha}\;.
\end{aligned}
$$

The class sizes are then obtained as

$$
\begin{aligned}
&\Delta G(b_k) := G(b_{k-1}) - G(b_k) = q^{(k-1)}(1-q)\;, \\
&\quad k = 1, 2, 3, \ldots
\end{aligned}
$$

We will just mention in passing that if α tends to infinity and we obtain an exponential distribution, then $q = (1 - 1/\alpha)^\alpha \to \mathrm{e}^{-1}$; with that $\Delta G(b_k)$ becomes independent of parameters. In bibliometric studies, we often observe α values in the range 2.0–3.0 (i. e., $q = 0.25$–0.30– [13.20, 21]). We show the corresponding class sizes around this value as well as for the exponential case in Table 13.1. For $\alpha = 3.0$, we have a theoretical distribution of objects over classes 1 through 4 roughly following a 70%–21%–6.5%–2.5% rule. Later on, in the empirical part, we will check how realistic this rule is in real-life situations.

Because of the underlying approach, i. e., the determination of classes on the basis of the characterization by conditional expectations, and the important role of the characteristic tail parameter α for the thresholds and class sizes, it was called the method of *Characteristic Scores and Scales* (CSS) when it was first proposed by *Glänzel* and *Schubert* [13.4]. Apart from

Table 13.1 Class sizes in relation to the α parameter of the Lomax distribution

α	q	Class 1 (%)	Class 2 (%)	Class 3 (%)	Class 4 (%)	Total (%)
∞	0.368	63.2	23.3	8.6	5.0	100
4.0	0.316	68.4	21.6	6.8	3.2	100
3.5	0.308	69.2	21.3	6.6	2.9	100
3.0	0.296	70.4	20.9	6.2	2.6	100
2.5	0.279	72.1	20.1	5.6	2.2	100
2.0	0.250	75.0	18.8	4.7	1.6	100

the above-mentioned role of the characteristic parameter, the knowledge of its actual value is not required for the method. The method does, therefore, meet the first requirement addressed in the introduction [13.4, p. 4]. Also, the second condition is satisfied, since the b_k thresholds are not preset, but are generated by the procedure.

So far we have merely dealt with the stochastic background of this model and its properties. In the following subsection, we will have a closer look at the statistical implementation of this theory.

13.2.2 Statistical Implementation of the Model

The statistical implementation of the above method of CSS is a recursive procedure of iteratively truncating a sample at its mean value, recalculating the mean of the truncated sample, and continuing this procedure until it is stopped, when the preset number of thresholds is reached or no new scores can be obtained. The following formalized description was provided in a previous study by *Glänzel* [13.20, pp. 93–94] and reads as follows:

We proceed from a bibliometric sample. This could, for instance, be a set of n papers published in a given subject field. The number of citations received by the individual papers in a given period are denoted by X_i, where index i runs from 1 to n. The observations are then ranked in descending order $X_1^* \geq X_2^* \geq \cdots \geq X_n^*$, with X_1^* denoting the citation rate of the most frequently cited paper and X_n^* consequently the lowest number of citations a paper within the set has received. We first put $b_0 := 0$ and $\upsilon_0 := n$, where n is the sample size, i. e., in the case of citation, the number of papers. According to the stochastic model of the previous section, b_1 is then defined as the sample mean

$$b_1 = \sum_{i=1}^{n} \frac{X_i}{n} = \sum_{i=1}^{n} \frac{X_i^*}{\upsilon_0} \, .$$

The value υ_1, which represents the size of the truncated sample, is uniquely defined by the following inequality

$$X_{\upsilon_1}^* \geq b_1 \quad \text{and} \quad X_{\upsilon_1+1}^* < b_1 \, .$$

This procedure can be repeated till it is stopped or the sample cannot be truncated further, so that we have

$$b_k = \sum_{i=1}^{\upsilon_{k-1}} \frac{X_i^*}{\upsilon_{k-1}} \, ,$$

and we choose υ_k so that

$$X_{\upsilon_k}^* \geq b_k \quad \text{and} \quad X_{\upsilon_k+1}^* < b_k, \quad k \geq 2 \, .$$

Obviously, the b_k thresholds define a not strictly increasing sequence, i. e., $b_0 \leq b_1 \leq b_2 \leq \ldots$, while the υ_k values (not strictly) decrease: $\upsilon_0 \geq \upsilon_1 \geq V_2 \geq \ldots$ If no preset number of classes is given, the procedure comes to a natural end if no further classes can be defined. In line with the intervals of the theoretical model, the k-th class is now defined by the pair of threshold values $[b_{k-1}, b_k)$ and the number of papers belonging to this class amounts to $\upsilon_{k-1} - \upsilon_k$. The share of papers falling into this class is the statistical equivalent of the theoretical value $\Delta G(b_k)$ (cf. Sect. 13.2.1).

On the basis of the above iteration, we can now define our citation-impact classes. Usually, we stop the iteration at $k = 2$ or 3. In what follows, we will use $k = 3$ as the default option. The interval $[b_0, \infty)$ contains the complete set $\upsilon_0 := n$. We call the first class $[b_0, b_1)$ with $(\upsilon_0 - \upsilon_1)$ elements the set of *poorly cited papers* since its elements are less cited than the average. The second class $[b_1, b_2)$ is referred to as *fairly* cited and contains $(\upsilon_1 - \upsilon_2)$ elements; the $(\upsilon_2 - \upsilon_3)$ papers of the third class $[b_2, b_3)$ are called *remarkably* cited and, finally, $[b_3, \infty)$ forms the group of the υ_3 *outstandingly* cited papers. Class 3 and 4 together can be considered highly cited. Combining those would, of course, be identical with the choice of $k = 2$.

After this step has been taken, we have the tools for scoring individual citation rates or any subsets or samples. Benchmarking exercises can now be readily constructed in the following manner. The original paper set, on which the above scores have been determined, serves as the reference distribution with a reduced state space. Now, observed citation rates of papers of any subset or sample can be gauged against this reference standard. The observed citation rate of each paper is compared to the thresholds b_k of the reference distribution and assigned to the four classes (or three classes, if k is set to 2), depending on whether the actual citation rate is greater or less than b_k. Note that the assignment is unique. Because of the definition of scores and the classes, multiple assignments cannot occur, not even for ties. This means that the third requirement formulated in the introduction in Sect. 13.1.2 is also met.

The choice $k = 2$ proved useful for small samples such as the publication output of individuals or research groups, whereas $k = 3$ is a general-purpose choice for large-scale analysis with sufficiently large publication sets in the samples under study.

When all elements of the subset or sample are each assigned to one of the classes, the two distributions can be compared. We will just mention in passing that the choice of $k = 1$ would result in *classical* bibliometrics with a comparison of mean values. Consequently, the last requirement in the list in Sect. 13.1.2, the seamless integration into the standard tools of scientometric performance assessment, is also met. The last question in the list concerning robustness and stability, along with its applicability to various levels of aggregation and its ability to gauge observations against several standards, will be answered on the basis of empirical evidence in the next section.

13.3 Characteristic Scores and Scales in Research Assessment

The method of CSS *includes* two basic properties that predestine it for universal use. Firstly, the CSS method is parameter-free, that is, subject to random errors of parameter estimators, and, secondly, it is self-adjusting. In other words, no predefined thresholds or quantiles are needed to set up the scales. In contrast, for instance, to percentiles, this approach is *natively* not sensitive to ties. The method can be interpreted as a reduction of the original citation distribution to a distribution over a given number of performance classes. The only arbitrary value that is needed to build the performance classes is its number. In practice, four classes are usually sufficient. As we will later see, at lower levels of aggregation, where the number of underlying publications is rather low, even three classes could be a more appropriate choice, since the fourth class might not contain enough elements. The four classes represent "poorly cited" (class 1), "fairly cited" (class 2), "remarkably cited" (class 3) and "outstandingly cited" (class 4) papers. Papers in classes 3 and 4 can be considered highly cited. Furthermore, the integration of the method into the standard scientometric tools directly results from its definition. The choice of two classes instead of four would provide a normalized mean citation score.

13.3.1 General Properties of the Method

The most striking advantage is the more detailed picture that we get regarding the citation impact. If the reference distribution is properly chosen, then CSS thresholds can be considered *normalized*. Unlike in the case of traditional indicators, where we can only infer whether the impact of our publication set under study is above or below the benchmark value, be it a mean or median, CSS provides five major types of deviations from the reference standard. And also here one can decide whether the possible deviation should be considered significant or not. Since we deal with distributions, the application of a simple χ^2 test is probably the most straightforward method for that, but alternatively, other or supplementary and preferably nonparametric statistics could be used, e. g., Cramer's V for the calculation of the effect size. The five paradigmatic types of deviations are symbolically potted in Fig. 13.1. Type III in the figure is in line with the reference standard. Two types, here marked as types II and IV, show clear trends towards poorly and highly cited papers, respectively. Type IV represents the most advantageous situation. Its distribution of the four performance classes is more skewed than the reference distribution: less poorly cited papers here entail more highly cited papers. Type II reflects the opposite situation: the large share of poorly cited papers is to the detriment of high-impact papers. These two types still correspond to their classical coun-

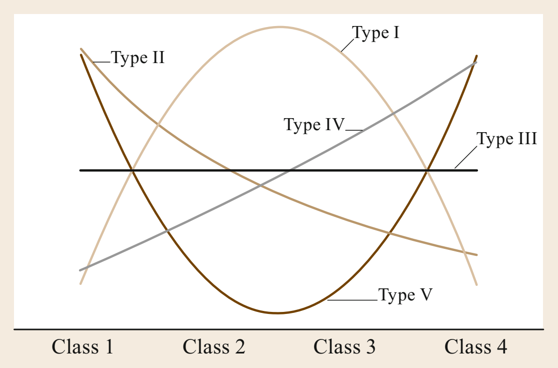


Fig. 13.1 The five different profiles according to their deviation from the reference standard (after *Glänzel* et al. [13.3])

terparts: high versus low (normalized) mean citation rates. Types I and II, however, represent completely different situations. Unlike type II, type I is characterized by the situation in which most papers are close to the average, and both poorly and highly cited papers are less frequent. Type V, finally, represents polarized citation impact. In such samples, poorly and highly cited papers are frequent and *average* papers are less frequent, as in the reference distribution. The latter two types are observable most notably at the level of individual scientists. The occurrence of these types thus depends on the level of aggregation, and the lower the level of aggregation, the more pronounced their shapes.

As already stressed by *Glänzel* et al. [13.3], a further advantage that this method shares with the percentiles approach is its insensitivity to outliers. An outlier is just one out of several publications that are assigned to the highest performance class. This prevents outliers from distorting the citation-impact indicators of the unit under study. As we mentioned in the methodological sections, the characteristic scores are obtained by iteratively calculating the mean value of a citation distribution and by subsequently truncating this distribution by removing all papers with fewer citations than the conditional mean. Although the method is based on mean values, the huge number of publications underlying the baseline, i. e., the complete database, guarantees that the influence of outliers remains marginal. From a statistical viewpoint, the publications that receive such striking citation rates form just a fraction of a thousandth of the population, i. e., the complete publication universe. The citation impact of this minuscule set of publications in the database that attracts this enormous amount of citations is easily absorbed by the total population. Nevertheless, these papers have been written and published by a rather small number of researchers, and if we assign the papers to statistics on these authors, their departments or even the institutes or organizations with which they are affiliated, the effect of the exceptional citation rates received by these papers can no longer be absorbed by the resulting samples. This was one of the main reasons why we replaced individual observations with performance classes. To give an example, we mention the case of the paper on the history of SHELX, which was published by a researcher at the University Göttingen. The extreme citation impact of this single paper is able to distort world university rankings if based on citation counts or means. *Waltman* et al. [13.7] first observed this effect. The paper in *Acta Crystallographica Section A* published in 2008 is still ranked first in the top list of the most cited papers during the last decade and is among the top ten most cited journal papers since 1955. We present the up-to-date citation impact of these ten most-cited papers just as an illustration (Table 13.2). Data have been sourced from Clarivate Analytics Web of Science Core Collection (WoS), but we stress that Elsevier's Scopus could be used as well. We will use the WoS throughout this chapter, so that we will explicitly indicate whenever an additional data source is referred to.

In addition to a review of our previous studies on CSS, the main objective of the remainder of the book chapter will be:

1. The demonstration of the applicability of the proposed method to various levels of aggregation, such as individual subjects on a large scale
2. The application to a combination of different subjects, including new results for journal assessment, and
3. The evaluation at the micro level.

The possibilities for applying the CSS method are manifold. In the following subsections, we will re-

Table 13.2 Extreme citation rates: the ten most-cited journal articles in the WoS displayed in descending order of citations received from publication year through 10.03.2017

1st author	Journal	Publication year	Volume	1st page	Citations
Laemmli, U.K.	Nature	1970	227	680	238 811
Bradford, M.M.	Analytical Biochemistry	1976	72	248	188 280
Sanger, F.	Proceedings of the National Adacemy of Science, USA	1977	74	5463	66 180
Chomczynski, P.	Analytical Biochemistry	1987	162	156	61 681
Becke, A.D.	Journal of Chemical Physics	1993	98	5648	60 685
Lee, C.T.	Physical Review B	1988	37	785	57 341
Perdew, J.P.	Physical Review Letters	1996	77	3865	54 938
Towbin, H.	Proceedings of the National Adacemy of Science, USA	1979	76	4350	54 353
Sheldrick, G.M.	Acta Crystallographica Section A	2008	64	112	52 956
Folch, J.	Journal of Biological Chemistry	1957	226	497	48 898

Data sourced from Clarivate Analytics Web of Science Core Collection

view them in a systematic way. The first group of applications comprises the assessment and profiling of publication sets according to their citation impact at various levels of aggregation. This concerns comparative analyses at the macro level (countries and regions), at the meso level, such as research organizations, universities, hospitals, university faculties and departments, and at the micro level, including research teams and individual scientists. A further opportunity provided by this method is the subject and journal analysis. We will just mention in passing that this is actually the field of application for which the method was originally designed [13.4]. Finally, the application to subject normalization was also soon discovered and implemented in a large compilation of journal–country indicators [13.22] and later used in the context of journal ranking [13.23]. The last option, which has not yet been frequently applied, is gauging citation classes against different reference standards that arise from several contexts. For instance, both the macro- and meso-level standard can be used to benchmark micro-level data, and at the same time, standard subjects or specifically delineated topics can be used to define the reference standards. A simplified overview of these fields of application is given in Fig. 13.2. The gray shapes on the left-hand side symbolize the most prominent domains of the assessment of research performance, while the three shapes in brown represent different aggregation levels of the underlying cognitive base. Although these three levels can also be studied independently of any national, institutional or individual research performance, even greater advantage can be gained if the results of the subject analysis are applied, for instance through improved subject/topic normalization, to research performance studies at different levels. In the following subsections, methodological and technical aspects will be discussed and illustrated by examples, where Fig. 13.2 will serve as a kind of road map throughout the chapter to navigate among the possible fields of application.

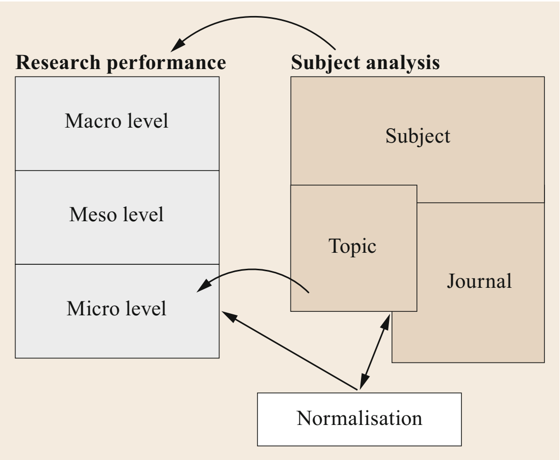


Fig. 13.2 Sketch of possible applications of Characteristic Scores and Scales

Before we explain in greater detail the methodological and technical aspects of the assessment of citation impact according to citation classes, we give empirical evidence of some important properties of CSS that are needed to properly interpret any results obtained from the application of this method. This concerns robustness with respect to subject matter, publication time and length of citation window. In the following subsection, we will give an illustration of these three aspects of robustness. Further evidence can be found in previous studies by the authors, notably in *Glänzel* [13.20, 23] and *Glänzel* et al. [13.24].

Finally, some words about the data sources. All calculations that have been made for the following (sub)sections are based on bibliographic data extracted from Clarivate Analytics Web of Science Core Collection. In order to study possible time dependence of CSS indicators, two volumes (2006 and 2013) of the WoS journal editions have been chosen. This choice allows for the calculation of citation measures for 3-year citation windows (i. e., 2006–2008 and 2013–2015, respectively) for both volumes and for a 5- and 10-year citation window for papers indexed for the 2006 volume (i. e., 2006–2010 and 2006–2015), since the last available volume for counting citations was 2015. Only *citable* papers, i. e., documents of type article, letter and review, have been selected for the chapter. All data were cleaned and processed for bibliometric use before the CSS indicators were built.

13.3.2 Empirical Properties and Disciplinary Characteristics of CSS

In order to shed light on subject-specific peculiarities, first, all documents extracted from the 2006 and 2013 volumes of the WoS database were assigned to the 74 individual subfields according to the modified Leuven–Budapest classification system (Fig. 13.3). *Glänzel* and *Schubert* [13.26] developed the first version of this scheme as a hierarchical system based on the Web of Science and Journal Citation Reports (JCR) subject categories. It was later modified [13.25] to provide a better categorization of the social sciences and humanities. In particular, the modification became necessary for the use with the *Book Citation Index* (BKCI) and the *Emerging Sources Citation Index* (ESCI) because of their larger coverage of the social sciences and humanities. Of course, this enhanced version is also fully compatible with the journal and proceedings editions,

THE LEUVEN-BUDAPEST CLASSIFICATION SCHEME FOR THE SCIENCES, SOCIAL SCIENCES AND HUMANITIES

0. MULTIDISCIPLINARY SCIENCES
- X0 multidisciplinary sciences

1. AGRICULTURE & ENVIRONMENT
- A1 agricultural science & technology
- A2 plant & soil science & technology
- A3 environmental science & technology
- A4 food & animal science & technology

2. BIOLOGY (ORGANISMIC & SUPRAORGANISMIC LEVEL)
- Z1 animal sciences
- Z2 aquatic sciences
- Z3 microbiology
- Z4 plant sciences
- Z5 pure & applied ecology
- Z6 veterinary sciences

3. BIOSCIENCES (GENERAL, CELLULAR & SUBCELLULAR BIOLOGY; GENETICS)
- B0 multidisciplinary biology
- B1 biochemistry/biophysics/molecular biology
- B2 cell biology
- B3 genetics & developmental biology

4. BIOMEDICAL RESEARCH
- R1 anatomy & pathology
- R2 biomaterials & bioengineering
- R3 experimental-laboratory medicine
- R4 pharmacology & toxicology
- R5 physiology

5. CLINICAL AND EXPERIMENTAL MEDICINE I (GENERAL & INTERNAL MEDICINE)
- I1 cardiovascular & respiratory medicine
- I2 endocrinology & metabolism
- I3 general & internal medicine
- I4 hematology & oncology
- I5 immunology

6. CLINICAL AND EXPERIMENTAL MEDICINE II (NON-INTERNAL MEDICINE SPECIALITIES)
- M1 age & gender related medicine
- M2 dentistry
- M3 dermatology/urogenital system
- M4 ophthalmology/otolaryngology
- M5 paramedicine
- M6 psychiatry & neurology
- M7 radiology & nuclear medicine
- M8 rheumatology/orthopedics
- M9 surgery

7. NEUROSCIENCE & BEHAVIOR
- N1 neurosciences & psychopharmacology
- N2 psychology & behavioral sciences

8. CHEMISTRY
- C0 multidisciplinary chemistry
- C1 analytical, inorganic & nuclear chemistry
- C2 applied chemistry & chemical engineering
- C3 organic & medicinal chemistry
- C4 physical chemistry
- C5 polymer science
- C6 materials science

9. PHYSICS
- P0 multidisciplinary physics
- P1 applied physics
- P2 atomic, molecular & chemical physics
- P3 classical physics
- P4 mathematical & theoretical physics
- P5 particle & nuclear physics
- P6 physics of solids, fluids and plasmas

10. GEOSCIENCES & SPACE SCIENCES
- G1 astronomy & astrophysics
- G2 geosciences & technology
- G3 hydrology/oceanography
- G4 meteorology/atmospheric & aerospace science & technology
- G5 mineralogy & petrology

11. ENGINEERING
- E1 computer science/information technology
- E2 electrical & electronic engineering
- E3 energy & fuels
- E4 general & traditional engineering

12. MATHEMATICS
- H1 applied mathematics
- H2 pure mathematics

13. SOCIAL SCIENCES I (GENERAL, REGIONAL & COMMUNITY ISSUES)
- Y1 education, media & information science
- Y2 sociology & anthropology
- Y3 community & social issues

14. SOCIAL SCIENCES II (ECONOMIC, POLITICAL & LEGAL SCIENCES)
- L1 business, economics, planning
- L2 political science & administration
- L3 law

15. ARTS & HUMANITIES
- K0 multidisciplinary journals
- K1 arts & design
- K2 architecture
- K3 history & archaeology
- K4 philosophy & religion
- K5 linguistics
- K6 literature

Fig. 13.3 The modified version of the Leuven–Budapest hierarchical classification scheme based on the WoS subject categories according to *Glänzel* et al. [13.25]

while major fields and subfields in the sciences of the previous version are not changed. The modified classification scheme is presented in Fig. 13.3.

The reason why we decided to use the second hierarchical level of the scheme results from experience. While major fields often prove too coarse—that is, subjects within the same major field might represent different standards in publication and citation behavior, for which mathematical & theoretical physics and particle & nuclear physics may serve as an example—

the choice of the lowest level, i. e., of the WoS subject categories, would result in a superposition of the multiple assignments of papers to subjects, which, to solve, requires an extensive combination of fractionating and weighting by subjects. Nevertheless, at the micro level, this would be—alternatively to the delineation of specific topics—the preferred solution, as lower levels of aggregation always need higher granularity.

Citation patterns are strongly influenced by subject characteristics. This effect is a well-known characteristic of citation indicators. Papers in mathematics, engineering and social sciences exhibit distinctly lower citation impact than their counterparts in the natural or life sciences. Even within major fields of the sciences and social sciences, communication patterns may vary considerably. *Peritz* [13.27] showed this in her study on the intradisciplinary differences in citation impact of theoretical, methodological and empirical papers in sociology in three prestigious sociology journals. In the sciences, theoretical subjects usually have a lower impact than experimental and applied ones. *Marx* and *Bornmann* [13.28] point to a further reason for subject-specific differences originating from the different coverage of the literature in multidisciplinary databases, which in turn is only partially a result of the scientists' distinct publication activities in the different disciplines. For these reasons, citation measures are—without proper subject normalization—not appropriate for cross-field comparisons. In order to illustrate this property, we show the aggregate impact factors (AIF) of 20 selected subject categories according to the 2015 sciences and social sciences edition of the Clarivate Analytics Journal Citation Report in Table 13.3. The selected categories represent the high end, the medium range and the low end of citation impact. The difference between the highest and the lowest impact extend to one order of magnitude.

As a consequence, the actual b_k values of the characteristic scores are also strongly influenced by the disciplines in which papers have been published. The subject dependence of characteristic scores, notably the enormous variation in the $b_3$values, has already been pointed out in earlier studies [13.20, 23, 24]. The underlying citation window is the second important factor that affects citation indicators, in general, and characteristic scores in particular. It goes without saying that larger citation windows provide the opportunity to attract and receive more citations than shorter ones. In order to illustrate both effects, we have fixed one publication year and selected a number of disciplines from different science and social science fields. The b_k values for the WoS 2006 volume and a range of subject fields in three different citation windows are given in Table 13.4. In particular, we have selected 18 of the 74 total subfields, two each from biosciences, internal and noninternal medicine, chemistry, physics, engineering and the social sciences, and one each from mathematics, geo- and space sciences, neuroscience and biology. The number of papers assigned to these disciplines 2006 ranges from roughly 10 000 to 60 000.

Table 13.3 Aggregate impact factors of 20 subject categories according to the 2015 JCR

Subject Category	AIF
CELL BIOLOGY	5.60
CHEMISTRY, MULTIDISCIPLINARY	5.59
NANOSCIENCE & NANOTECHNOLOGY	5.55
MULTIDISCIPLINARY SCIENCES	4.90
CELL & TISSUE ENGINEERING	4.70
CHEMISTRY, PHYSICAL	4.64
HEMATOLOGY	4.40
⋮	⋮
OPTICS	2.22
CRYSTALLOGRAPHY	2.22
WATER RESOURCES	2.21
PSYCHOLOGY, MULTIDISCIPLINARY	2.20
SOIL SCIENCE	2.19
MEDICAL LABORATORY TECHNOLOGY	2.17
⋮	⋮
EDUCATION & EDUCATIONAL RESEARCH	1.06
SOCIAL WORK	1.06
ENGINEERING, MARINE	1.04
ENGINEERING, AEROSPACE	1.02
LINGUISTICS	1.01
HISTORY & PHILOSOPHY OF SCIENCE	0.84
MATHEMATICS	0.74

Data sourced from Clarivate Analytics Journal Citation Report

The b_k values ($k = 1, 2, 3$) are arranged by citation window in three columns each (Table 13.4). The first three columns corresponding to citations received in 2006–2008 are completely in line with the AIF presented in Table 13.3, and thus confirm the general trend of subject-specific characteristics in citation indicators, although the AIF is a synchronous indicator with fixed citation year and variable publication years. Mathematics and social sciences are found at the low end, while biosciences, notably cell biology, are at the top surrounding the natural sciences, medical sciences and psychology. The high and the low ends are separated by one order of magnitude, as in the case of CSS. Increasing the citation window yields considerable changes in all b_k values. The increase by 2 years (5-year window 2006–2010) results in roughly doubling of the corresponding values. Further enlarging the citation window to 10 years (2006–2015) yields another doubling. This trend is proportional. Although this is not important for

Table 13.4 Characteristic scores for 18 selected subfields in 2006 according to the Leuven–Budapest scheme

Subfield	Papers	3-year citation window			5-year citation window			10-year citation window		
		b_1	b_2	b_3	b_1	b_2	b_3	b_1	b_2	b_3
B1	60 673	8.23	19.43	36.18	16.55	38.40	73.28	34.32	84.75	174.21
B2	21 220	11.61	28.22	54.28	23.47	58.11	114.19	48.63	127.03	266.10
C1	32 443	4.72	10.57	18.46	9.35	21.62	37.85	18.89	44.56	83.33
C3	25 781	5.37	12.04	21.05	10.70	22.99	38.78	21.83	48.47	87.84
E2	37 378	3.03	8.74	15.99	5.95	15.07	30.06	13.61	39.13	87.82
E3	17 026	2.56	6.71	12.06	5.97	15.32	29.70	15.33	44.82	91.00
G1	15 585	8.24	20.72	37.66	14.82	36.65	68.92	27.44	71.68	142.54
H2	16 717	1.38	3.89	6.43	3.04	8.44	15.61	6.91	18.29	38.51
I1	26 360	6.39	16.83	31.77	12.88	32.96	63.52	26.24	70.83	142.09
I5	20 393	8.08	20.67	39.77	15.79	38.35	76.06	31.13	79.29	160.58
L1	18 596	2.08	6.09	11.32	5.78	14.57	26.93	17.80	49.60	99.57
M4	12 597	3.26	8.56	15.88	6.98	16.37	30.57	15.01	37.37	69.13
M6	30 380	5.35	13.71	24.95	11.61	28.37	53.23	25.84	64.73	124.83
N2	26 694	4.05	10.20	18.27	9.81	22.21	40.31	25.59	62.98	121.10
P2	14 714	5.01	11.55	19.20	9.63	21.14	37.82	19.34	46.60	90.70
P5	12 953	5.71	15.58	30.86	9.77	26.65	52.64	17.19	49.22	100.15
Y2	13 489	2.44	6.29	11.13	6.14	15.15	26.25	16.14	40.20	72.63
Z3	39 906	6.18	14.77	26.13	12.82	29.06	54.30	27.88	67.09	135.12

Data sourced from Clarivate Analytics Web of Science Core Collection

any field of application of CSS scores, it reveals an interesting empirical property. A simple linear regression analysis shows a very strong correlation with r^2 values ranging from 0.97 to 0.99 for all citation windows and k values, while the slope varies from 2.0 to 2.1 for the extension of the citation window from 3 to 5 years, and takes values between 2.1 and 2.4 for the extension by an additional 5 years. Although there is no theoretical model or rationale that could explain these values, from an application viewpoint, this simple factor of around 2 could potentially be used for time-based normalization of citation indicators.

In contrast to the characteristic scores, the distribution of citations over the four classes that are defined by the b_k scores is strikingly insensitive to both the citation window and the underlying discipline. This kind of robustness is in line with the observations published in the above-mentioned studies by *Glänzel* [13.20, 23] and *Glänzel* et al. [13.24]. This stability is maintained even if the citation window is further extended to 20 years, on the basis of the 1980 volume of the Science Citation Index (SCI) and a 21-year citation window. The share of those papers that have received at least b_k but not reached b_{k+1} citations (with $b_0 = 0$ and $b_4 = \infty$) for all papers published in each of the 74 Leuven–Budapest subfields in 2006 was calculated for the three periods 2006–2008, 2006–2010 and 2006–2015. Table 13.5 gives the corresponding values for the same selection of disciplines as above. The class-percentage values are very similar, although not identical, independent of the discipline and the citation window. The conclusion drawn from these empirical distributions is completely in line with both our Paretian model (Sect. 13.2.1) and the empirical observations made by, e. g., *Albarrán* and *Ruiz-Castillo* [13.29] and *Glänzel* et al. [13.24]. According to these results, the share of papers cited less frequently than the average (class 1) amounts to roughly 70%, the share of those assigned to class 2 to about 21%, and the share of papers in the upper two classes is 6%–7% and 2%–3% of all papers, respectively.

Apart from just showing empirical properties of CSS thresholds and classes, there is also a straightforward application of these results. The shares of papers falling into classes 1 through 4 can be used as a benchmark for any sample drawn from a population or for any other subset of papers in the corresponding subfield. The three corresponding thresholds (b_k scores) then serve as the criteria of assignment for the papers in the subset under study. If this subset were the true mirror of the entire population, its share in all four classes would be identical to that of the reference standard. Any deviation from this standard indicates a specific profile. Possible paradigmatic profiles have already been sketched in Fig. 13.1 (Sect. 13.3.1). In particular, individual profiles might be more or less skewed with higher or lower shares in the lower classes, respectively, or more or less polarized as the lower/higher share of lower-class papers is compensated by a higher/lower share of upper-class papers. This makes comparison with benchmark values or reference stan-

Table 13.5 CSS-class percentage shares for 18 selected subfields in 2006 according to the Leuven–Budapest scheme

Subfield	3-year citation window				5-year citation window				10-year citation window			
	Class 1	Class 2	Class 3	Class 4	Class 1	Class 2	Class 3	Class 4	Class 1	Class 2	Class 3	Class 4
B1	69.7	21.5	6.3	2.5	69.3	21.9	6.3	2.4	71.3	21.0	5.6	2.0
B2	69.9	21.5	6.0	2.6	70.9	20.9	5.8	2.4	72.3	20.2	5.5	1.9
C1	65.1	23.7	7.7	3.5	67.9	21.9	7.0	3.1	68.2	22.4	6.7	2.7
C3	67.0	23.2	7.0	2.9	65.8	22.9	7.8	3.5	67.4	22.9	6.9	2.8
E2	73.1	18.4	5.8	2.6	67.9	23.1	6.6	2.4	72.8	20.3	5.2	1.7
E3	68.5	21.2	7.5	2.8	68.4	22.2	6.7	2.7	73.1	19.2	5.7	2.1
G1	69.7	20.8	6.6	2.8	69.0	21.6	6.6	2.7	71.0	20.7	6.0	2.3
H2	70.4	18.5	7.8	3.3	72.6	19.3	5.7	2.4	70.1	21.9	5.9	2.1
I1	70.5	20.5	6.4	2.6	69.6	21.3	6.5	2.6	71.6	20.2	5.9	2.3
I5	71.8	20.1	5.8	2.4	69.7	21.8	6.0	2.4	71.6	20.6	5.6	2.2
L1	73.4	19.1	5.2	2.3	68.5	21.7	6.6	3.2	71.8	20.0	5.7	2.5
M4	70.4	20.8	6.2	2.6	66.5	23.5	7.1	2.9	70.1	21.0	6.6	2.3
M6	69.9	20.7	6.4	3.0	68.7	22.1	6.4	2.8	69.5	21.6	6.3	2.6
N2	70.6	20.8	6.1	2.5	66.6	23.4	7.1	2.8	69.9	21.4	6.3	2.5
P2	69.4	20.7	7.1	2.8	66.7	23.4	6.9	2.9	70.2	21.4	6.3	2.1
P5	70.8	20.6	6.1	2.5	71.2	20.2	6.1	2.5	73.1	19.2	5.7	2.0
Y2	68.3	21.8	7.1	2.7	68.6	21.6	6.9	3.0	69.1	21.4	6.6	2.9
Z3	69.5	21.0	6.8	2.8	67.5	23.1	6.6	2.8	69.8	22.0	5.9	2.2

Data sourced from Clarivate Analytics Web of Science Core Collection

dards more complex than the usual higher/lower than expected as known from traditional bibliometric exercises. In order to infer significance of deviation from the reference standard, a simple χ^2 homogeneity test with three degrees of freedom can be applied. In this context, it is also important that the sample size is large enough, that is, it should amount to at least 20–30 items (*Vincze* [13.30] mentions a minimum of 40 items for the Welch test, which requires slightly larger sample sizes). In order to illustrate how to use CSS for sample profiling, we give the following example.

We have taken four nonrandom subsets of publications, one each for cell biology (B2), psychiatry & neurology (M6), analytical, inorganic & nuclear chemistry (C1) and pure mathematics (H2). The papers were published in 2006, while the citation window is 2006–2008. The sample sizes of these subsets range from 100 to 329 and are thus sufficiently large. Applying the χ^2 homogeneity test, we obtain $\chi^2 = 7.99$ for cell biology, $\chi^2 = 38.30$ for psychiatry & neurology, $\chi^2 = 5.23$ for the chemistry discipline and $\chi^2 = 4.32$ for mathematics. The critical value at a confidence level of 95% is 7.81. According to these results and the CSS-class distributions of the samples and the reference standard, we can conclude that the distribution of papers over CSS classes of the first sample (B2) differs significantly from that of the benchmark, and thus follows profile type II. This type reflects a quite unfavorable situation, with more less-cited papers and fewer papers in the upper classes. The deviation from the reference standard is even more striking in the second sample (M6), but here the deviation points in another direction. The share of poorly cited papers lies distinctly below the standard, while the percentage of fairly and highly cited papers in all classes is above the corresponding benchmark value. This situation corresponds to type IV, which reflects the most advantageous situation. The paper set in analytical, inorganic & nuclear chemistry (C1) would be an excellent example of type I. However, the deviation from the standard is not significant. Finally, the fourth subset (H2), because of its relatively small size, also does not differ significantly from the standard. Its CSS profile by and large follows the benchmark distribution and is of type III. The four distributions with the χ^2 test values are displayed in Table 13.6.

13.3.3 Application of Characteristic Scores and Scales in Comparative Macro- and Meso-Level Studies

Using a dataset reflecting a properly delineated homogeneous subject profile as the reference standard is, of course, one of the possible applications for the impact analysis of research output, and we will, therefore, deepen this further in this section. However, the typical task in quantitative research evaluation is the assessment of rather heterogeneous publication sets, such as the multidisciplinary output of a research unit or a country. It is, of course, possible to process each discipline separately and assess research performance in each dis-

Table 13.6 Four samples to illustrate benchmarking according to CSS classes (publication year: 2006, citation window: 2006–2008)

Subfield	Subfield reference standard				Subset					χ^2
	Class 1	Class 2	Class 3	Class 4	Class 1	Class 2	Class 3	Class 4	N	
B2	69.9	21.5	6.0	2.6	79.6	17.5	2.2	0.7	137	7.99
M6	69.9	20.7	6.4	3.0	54.1	31.3	10.0	4.6	329	38.30
C1	65.1	23.7	7.7	3.5	63.0	30.0	7.0	0.0	100	5.23
H2	70.4	18.5	7.8	3.3	69.3	16.7	11.4	2.6	228	4.32

Data sourced from Clarivate Analytics Web of Science Core Collection

cipline, but real-life situations relate rather to output of inter- and multidisciplinary research results, which preferably should not be disaggregated into individual subjects. Referring to Fig. 13.2, the subject of analysis of the previous subsection was located at the upper-right corner of the diagram. We now move to the left-hand side, where macro- and meso-level publication output is usually not assignable to single disciplines and thus requires more sophisticated approaches. *Glänzel* et al. have described such an approach [13.24].

Even if CSS are applied in multidisciplinary environments, all publications that are subject of the analysis need to be assigned to individual disciplines. The reason is that the *unique* assignment of publications to performance classes is an indispensable precondition for the applicability of CSS. We illustrate this problem using the following example. Bibliographic databases allow and apply multiple assignments of indexed items to subject classification, independent of whether the assignment is made on papers directly (e. g., MathSciNet) or through the journal in which it is published (e. g., WoS, Scopus). Let us now assume that a paper is assigned to two subjects, here denoted by S1 and S2. Both subjects usually have their own specific publication and citation standards. As we could observe in Table 13.4, the two subjects might have distinctly different b_k ($k = 1, 2, 3$) thresholds. Let us further assume, without loss of generality, that the citation standard of S1 is higher, that is, its b_k values are higher than their counterparts of S2. This means that the paper in question might be assigned to class 3 in subject S1 and to class 4 in S2, since its citation rate does not exceed b_3 in S1, but it is greater than the corresponding threshold b_3 in S2. Directly combining the two subjects or applying a simple average of b_k would not solve this problem or provide any acceptable solution, as this solution would lack any theoretical background or justification. We have solved this problem by applying a proper subject-based fractionation such that each publication is gauged against only one individual standard and b_k threshold value for each $k = 1, 2, 3$. The basic idea behind this solution was developed in the context of traditional indicators. Unlike the journal-based expected citations rates of individual papers, discipline-based expectations are subject to the same methodological problem of setting unique reference standards. As argued in a previous study by *Glänzel* et al. [13.31], one consequence of multiple assignments is the necessity of fractionation by all individual subjects to which the paper is assigned, and this has to be followed by the calculation of proper weights for the corresponding individual subject-expected citation rates. This must be done by defining correct weights such that the sum of the individual subject-based expectations over all publications in the system equals the citation total of the database in the combination of these fields. As a result, one obtains an "implicit" classification with standards that are different from those given in Table 13.4, since the corresponding thresholds are influenced by the combination of individual co-assignments to disciplines. As a consequence, we do not have *common* thresholds for all papers assigned to the same discipline. In order to distinguish the two types of scores, we denote the common scores in subsection 3.1 by b_k and the individual ones according to this procedure by b_k^*.

Also, this procedure is iterated to determine all CSS scores needed to define the four classes. This is done in the following manner. The first step is identical to the procedure for calculating subfield-expected citation rates [13.31]. A first fractionation applies to the calculation of the citation means of the disciplines. This is done on the basis of the respective number of subfields to which a publication is assigned. In this step, publications as well as citations are fractionated. After this is done, the individual expectation denoted by b_1^* is calculated for each paper, which is simply the mean value of the fractionated subfield standards. As we did this before in the disciplinary approach in Sect. 13.3.1, all papers that received fewer citations than their individual expectation are removed from further calculation. The above procedure is then repeated on the remaining set. This way we obtain the individual thresholds b_2^*. In total, the procedure is applied three times to obtain the three individual characteristics scores ($k = 1, 2, 3$) for each paper. After this is finished, papers can be uniquely assigned to one of the four classes. We would like to stress that this model is an extension of the previous single-disciplinary model. In particular, if the underlying paper set consists only of publications from a single discipline, fractionation and weighting would result in a trivial solution that would be identical to the unfractionated solution described in the previous subsection. The individual b_k^* thresholds in this case are identical to the common characteristic scores b_k.

In order to illustrate this procedure at the macro and meso levels, we have first selected 25 countries from among all regions in the world. We have selected two publication years, 2006 and 2013. The choice of the first year (2006) allows the calculation of indicators, i. e., of CSS scores, and the corresponding classes for citation windows up to 10 years. The second publication year allows for a comparison of the values of the same indicators shifted over 7 years to monitor their evolution. When selecting the countries, we aimed at both sufficiently large publication sets underlying the statistics in the two publication years and an acceptable geographic and geopolitical coverage. We have included papers of document type article, letter and review published in any subject category. The authors have published similar results by using the publication years 2007 and 2009 with a 5- and 3-year citation window, respectively [13.24]. This also gives us the opportunity to compare part of the new results with the previous, but differently structured results, since the overlap of the country selection is quite large. This comparison could confirm or argue against the robustness of the method.

First, we present the distribution of papers over CSS classes for the 25 selected countries. We have used the three-letter ISO codes for the country names; the key can be found in the Appendix. Table 13.7 gives the percentages along with the size of the underlying dataset. The last row displays the reference standard. According to the results published in our previous study, the reference standards for 2007 with a 5-year citation window and 2009 with a 3-year citation window amounted to 69.8% (class 1), 21.5% (class 2), 6.3% (class 3) and 2.4% (class 4), and 69.7% (class 1), 21.4% (class 2), 6.4% (class 3) and 2.5% (class 4) [13.24]. The

Table 13.7 Distribution of national shares of 25 countries' publications over the reference CSS classes in 2006 in all fields combined using two different citation windows (in alphabetical order)

Country	Papers	3-year citation window				10-year citation window			
		Class 1 (%)	Class 2 (%)	Class 3 (%)	Class 4 (%)	Class 1 (%)	Class 2 (%)	Class 3 (%)	Class 4 (%)
AUS	31 028	63.9	24.9	7.8	3.3	65.2	24.7	7.2	3.0
BEL	14 135	60.9	26.0	9.0	4.0	63.6	25.3	7.9	3.3
BRA	19 976	78.6	16.3	3.7	1.4	78.4	16.9	3.5	1.2
CHE	18 622	56.1	28.2	10.4	5.2	59.6	27.1	9.2	4.1
CHN	84 312	74.6	18.3	5.1	2.0	75.9	17.5	4.9	1.7
CZE	6639	71.0	20.8	5.8	2.4	75.4	18.6	4.3	1.7
DEU	80 177	62.9	25.1	8.4	3.6	66.8	23.5	7.0	2.6
DNK	9668	57.2	27.9	9.7	5.2	59.1	28.4	8.4	4.0
ESP	35 335	65.9	24.0	7.2	2.8	68.9	22.8	6.2	2.2
FRA	57 281	65.1	24.0	7.8	3.2	67.6	23.3	6.7	2.4
GBR	87 695	62.6	25.1	8.5	3.9	64.4	24.6	7.7	3.2
IND	28 978	78.8	16.2	3.7	1.2	78.7	16.3	3.7	1.2
ISR	11 641	66.4	23.4	7.0	3.1	68.8	22.5	6.2	2.4
ITA	44 756	65.9	24.0	7.1	2.9	69.2	22.6	6.0	2.2
JPN	78 236	72.1	20.6	5.3	2.0	76.3	17.9	4.3	1.4
KOR	28 799	74.8	18.9	4.7	1.6	77.6	17.2	4.0	1.2
NLD	25 932	57.4	28.2	9.8	4.6	59.9	27.4	8.7	3.9
POL	15 248	77.1	17.4	4.0	1.6	81.1	14.7	3.1	1.2
RUS	22 451	84.1	11.8	3.0	1.1	87.0	10.1	2.2	0.7
SGP	7110	64.1	24.8	7.9	3.3	66.5	23.8	7.3	2.4
SWE	17 842	61.6	26.5	8.1	3.7	63.9	25.6	7.5	3.0
TUR	15 839	81.4	14.3	3.1	1.2	80.5	14.9	3.5	1.1
TWN	18 363	72.0	21.1	5.2	1.8	73.4	20.7	4.4	1.4
USA	323 420	60.6	25.9	9.1	4.3	62.6	25.5	8.2	3.6
ZAF	5547	72.5	19.4	5.6	2.5	73.9	19.2	5.0	1.8
TOT	1 059 046	69.9	21.2	6.3	2.6	71.9	20.4	5.6	2.1

Data sourced from Clarivate Analytics Web of Science Core Collection

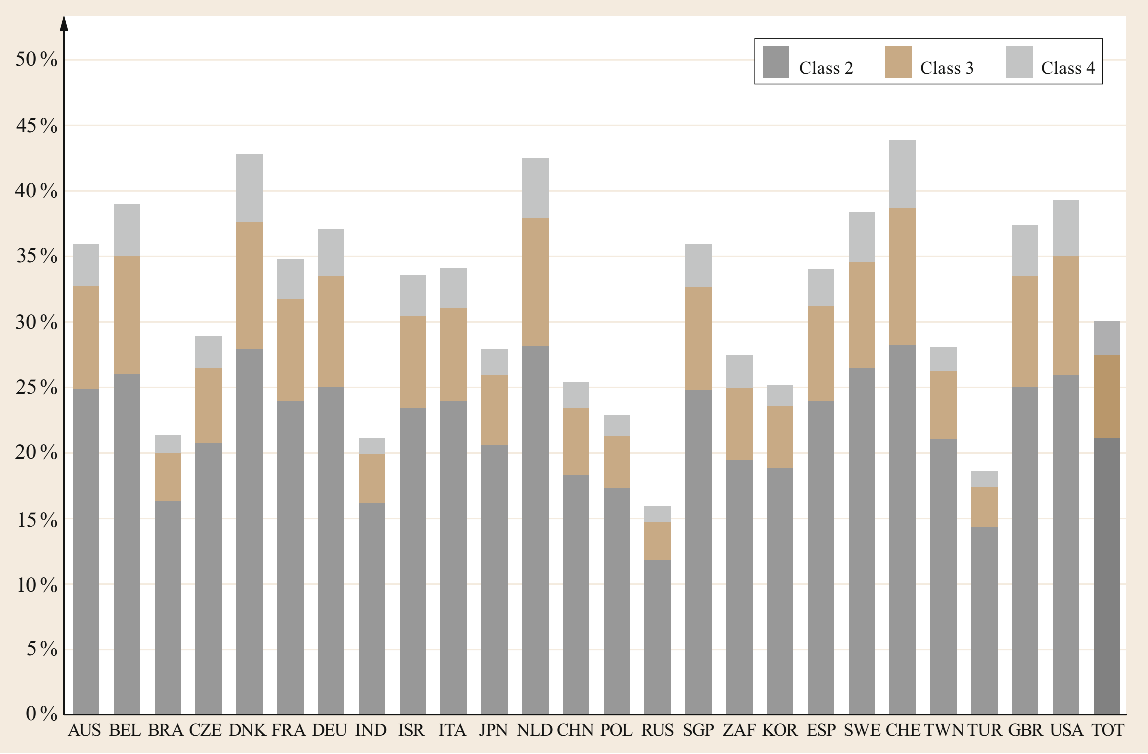


Fig. 13.4 Distribution of national shares of 25 countries' publications over the upper reference CSS classes in 2006 in all fields combined (citation window: 2006–2008) (Data sourced from Clarivate Analytics Web of Science Core Collection)

present results are very similar, although we can observe a slightly more skewed situation with a 10-year citation window. Comparing these distributions, most notably those for 3- and 5-year citations windows, we can again confirm the overall robustness of the method, with a class distribution of roughly 70%–21%–9% (class 1, class 2 and class 3&4), which has also been confirmed by *Albarrán* and *Ruiz-Castillo* [13.29]. Even more interesting, the CSS distribution of the selected countries shows the same robustness, while national characteristics become obvious. At the macro level, one cannot expect very polarized patterns (type I or V according to Fig. 13.1), and indeed, most shapes follow types II, IV or roughly type III. A χ^2-test is omitted here since, as a consequence of the large datasets at this level, almost all small deviations from the benchmark are to be considered significant. The strikingly large share of highly cited papers in Switzerland, Denmark, the Netherlands, the USA and Belgium has already been mentioned by *Glänzel* et al. [13.24]. The CSS classes of these countries are the most striking examples of type IV distributions. The type IV distribution of Singapore is not so pronounced as in the case of the above-mentioned countries, but is somewhat surprising. This is contrasted with the clear type II shapes of several countries including Russia, Brazil, India, China and Poland, but Japan and South Korea also belong to this group. Among the larger Eastern European countries, only the Czech Republic and Hungary (not displayed in Table 13.7) are close to the standard type III. Their patterns remain stable if the citation window is enlarged.

The following example shows CSS distributions of the same country sets for papers published in 2006 and 2013, each with a 3-year citation window. This underpins the comparison of the same indicators over a temporal distance of 9 years. Instead of the presentation of indicators in tables, we chose a visualization in stacked bars. Since percentages summed up over the four classes give 100%, we have omitted the lowest class. We must stress that a similar share of class 1 papers does not imply the same distribution over the upper classes. Figures 13.4 and 13.5 show the distribution over classes 2–4 of the 25 countries, along with the corresponding reference standard. The bar at the right-hand side represents the world standard. The results are very clear, although a detailed interpretation of the bars is not straightforward.

Essential changes cannot be observed, but those were not expected. China, Korea and India have gained citation impact; most notably, Singapore's share of

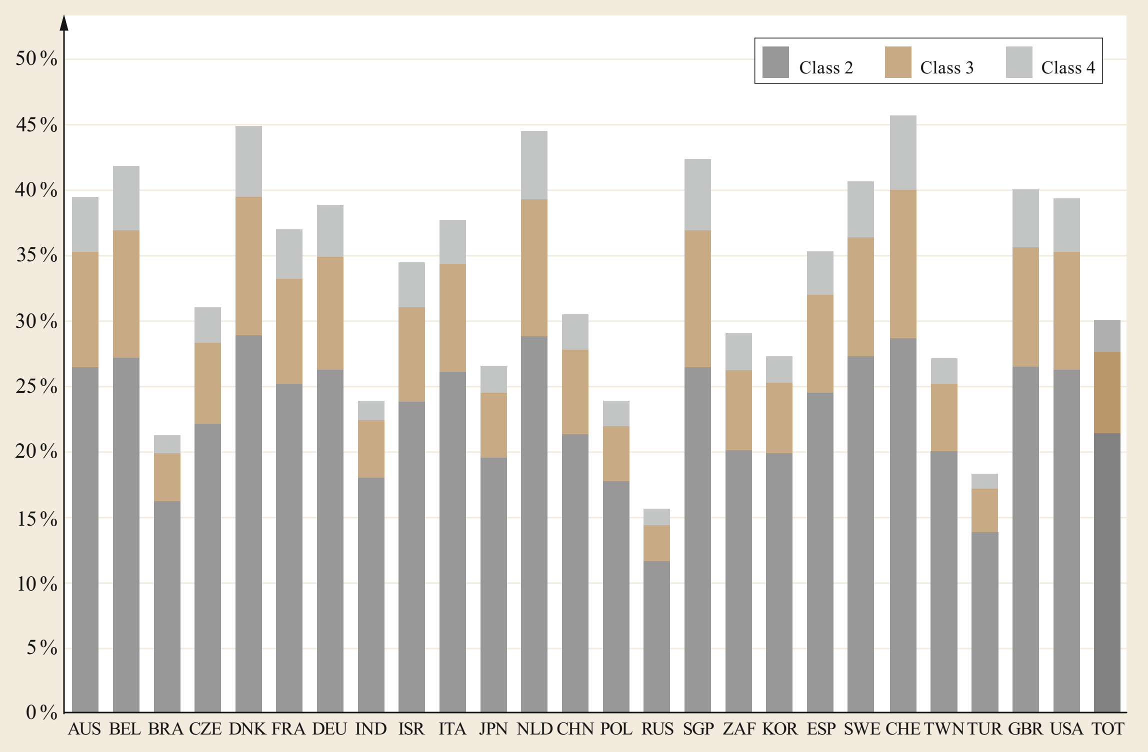


Fig. 13.5 Distribution of national shares of 25 selected countries in the upper reference CSS classes in 2013 in all fields combined (citation window: 2013–2015) (Data sourced from Clarivate Analytics Web of Science Core Collection)

fairly and highly cited papers has grown considerably, reaching the level of the most developed European and North American countries in 2013. By contrast, Russia, Brazil, Taiwan and Japan slightly lost impact during these 9 years. The overall picture obtained reflects stability.

To conclude this subsection, we illustrate the applicability of CSS scoring to the institutional level. For this purpose we have selected two universities each from 10 European countries. Here we have two reference standards, the national one for each country and the benchmark on the basis of the complete population. Of course, one would expect that the universities' profiles should mostly mirror the national patterns, but we can also find situations that differ from the respective national standard. These might be more or less favorable than the corresponding national profile.

Figure 13.6 shows the university profiles along with their national and the world reference standards for 2013, with a 3-year citation window. The respective Belgian, German, Italian, Swedish and British universities reflect slightly or even distinctly more favorable profiles than their national reference standard. In the other sample countries we also find universities with less favorable profiles. In addition to the large multidisciplinary universities, we have selected large medical (HU1 and SE2) and technical universities (DK2, DE2 and NL2). The high standard of technical universities again provides evidence of the subject independence of the method, since the applied and technical sciences generally attract fewer citations than the natural sciences, most notably the life sciences. Further examples can be found in the aforementioned study by the authors [13.24].

13.3.4 Application of Characteristic Scores and Scales to Micro-Level Studies

While bibliometric macro- and meso-level studies generally represent quite similar situations in terms of subject profiles (multidisciplinarity) and sample sizes (statistical reliability), studies of research teams face two specific issues. The first concerns the specific research and publication profiles of individual scientists or research teams. These profiles often cannot be captured by the standard subject categories of disciplines. Here, specialization and interdisciplinarity are the predominant models, and their fields of activities might be characterized by distinctly different citation behaviors from the standard subjects. This requires a proper

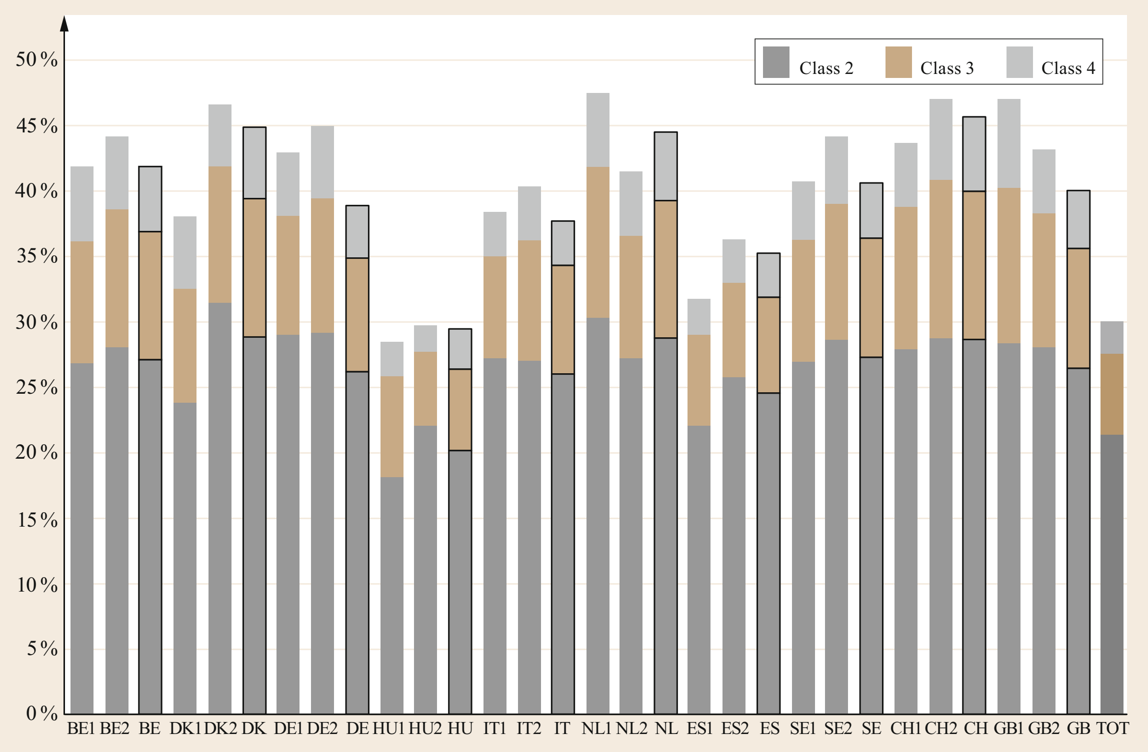


Fig. 13.6 Distribution of shares of publications of ten selected universities and countries in all fields combined over the upper reference CSS classes in 2013 (citation window: 2013–2015) (Data sourced from Clarivate Analytics Web of Science Core Collection)

subject delineation and determination of benchmarks according to the researchers' profiles. According to Fig. 13.2, we have to go back to journals and research topics in order to find the appropriate normalization for CSS benchmarking at this level of aggregation. The other issue is a consequence of the small sample sizes. Not all researchers (or teams) active in the field under study have published a sufficient number of papers to be included in the comparison group, and more prolific scientists (or teams) generally achieve higher visibility and impact than single-paper authors or less active scientists. This may result in built-in biases if the publication sample for individuals originates from the output of the more active authors [13.32]. In other words, the selection of topics and reference groups might determine different reference standards and baseline distributions for the publication set under study. Furthermore, at this level, outliers and observations differ distinctly from their expectation and distort traditional indicators to a greater extent. That is why micro-level studies require extremely stable and robust solutions. Again, we will show that the CSS-based method meets these requirements. As with any application of bibliometric methods to the micro level, i. e., to the level of individual scientists and research teams, the CS remains a challenging task [13.33]. According to multiple evidence-based statements, evaluation criteria should never rely on quantitative methods alone and should always be combined with appropriate qualitative methods. Furthermore, if the underlying data set is not sufficiently large, quantitative methods should not be applied at all.

The selection of a proper benchmark is another challenge. While we can use a sufficiently fine-grained subject classification at the meso and macro levels, the research assessment of individuals and teams requires personalized solutions. Typically, we have two different approaches at this level, the bottom-up and the top-down. The top-down approach proceeds from a given topic which is largely different from any predefined substructure of the subject-classification schemes that come with or are derived from bibliographic databases. Then, one has to identify the actors, i. e., authors and/or research teams who are active in this topic. Finally, indicators—in our case CSS scores—must be calculated for the topic and the authors' citation distributions gauged against this standard. One drawback of this solution is that only those papers of relevant authors/teams that fall within the topic are taken into account. If authors are also active outside this topic, that part is thus ignored. This approach, which represents a topic–author (topic–team) combination, is nevertheless the easier way compared with the bottom-up approach, which has not always been a unique solution. The authors are given, and their research topics must be delineated on the basis of their publication activity. The broadness of the subjects, be it ready-made subject categories or individually delineated tailor-made topics, is normally determined in collaboration with the agency or authority that commissions the bibliometric study. Preserving the authors' anonymity, we chose the first option (top-down approach) to illustrate the application of the CSS method to individual scientists. We have chosen the topic *scientometrics*, a subject we are very familiar with. Since we also know the actors, we can readily judge the correctness of inference from the data. We have collected publication data from 1998 to 2012 and applied 3-year citation windows, shifted by publication year. This citation window is admittedly short, but permits the inclusion of quite recent papers, which is an important aspect for a growing topic like scientometrics. In order to delineate the field, we have used core journals, including *Scientometrics* and *Journal of Informetrics*, along with a combination of a large range of keywords, including sciento-/biblio-/infor/web(o)-metric*, (co-)citation analysis*, bibliographic coupling, co-authorship and network/analysis. The total number of relevant papers amounts to 11 514. Smaller disciplines or specialized topics that form a subpart of a discipline, or the papers that are assigned across several disciplines but not fully covered by those, usually produce biased citation-impact baselines, if standard subject classification is used [13.34].

In the case of scientometrics, the argument against the use of standard fields is that papers in this topic exhibit, on average, higher citation rates than other papers in information and library science. Many relevant papers are published in journals outside this subject category, notably in multidisciplinary and specialized journals, practically in all fields of the sciences and social sciences. However, if we determine the CSS reference distribution of the citations received by the bibliometric papers and ignore any subject assignment, we obtain a class distribution of 67.6%–23.1%–6.4%–2.9%, which is again in line with the general distribution rule. This again substantiates robustness, most notably if one takes the large publication period of 15 years into account. If we exclude authors with fewer than ten papers each, we obtain a distribution of 54.0% (class 1), 31.1% (class 2), 9.7% (class 3) and 5.3% (class 4), which again illustrates the strong bias towards prolific authors [13.32]. We also see that the profiles of the most prolific authors, i. e., of the 22 authors with at least 30 papers each, tend to have profiles according to type IV. We have assigned identifiers to all authors who appeared as co-authors in publications collected for the scientometrics dataset. Only one author among the 22 persons displayed in Table 13.8 has a type I profile, due to the large share of fairly cited papers (#654). Authors whose profiles proved not to deviate significantly from the reference standard are automatically considered to have a type III profile.

Two issues need to be mentioned in the context of inference from CSS statistics at this author-level aggregation. First, authors with about 40 papers need only a single class 4 paper to get in line with the expectation, since this amounts to $\approx 2.5\%$, and second, co-authorship has a strong effect on the high end of the CSS class distribution here. This is the reason we recommend the application of three CSS classes (with $k = 1, 2$) only. The five profile types are still distinguishable using three classes. Alternatively, a χ^2 test could be applied with two degrees of freedom and the corresponding critical value of 5.99. In this context, we again must stress the necessity of context analysis required at the micro level, which also includes seniority and career analysis, citation context and analysis of co-authorship networks [13.33, 35].

13.3.5 Application of Characteristic Scores and Scales to Journal Assessment

Moving further to the right-hand side of Fig. 13.2, we arrive at the journal area. From a historical viewpoint, the use of the CSS in the context of journal assessment was the first application, and this was actually the application for which the method was designed [13.4, 22, 36]. Unlike subject assignment or the profiles of individuals and research units or countries, journals do not require multiple assignments of papers. Journals form a true partition of the document space that is covered by a bibliographic database and also of any paper set under study. If papers are to be assigned to journals on the basis of where they have been published, assignment is always unique, and thus no fractionation is needed. This property essentially simplifies the application of bibliometrics to journal indicators. On the other hand, journal analysis is still an important methodological topic in scientometric research and an indispensable fundamental for supplementing bibliographic databases by journal metrics (cf. Journal Citation Reports (JCR), Scimago Journal Ranking (SJR), Scopus CiteScore metrics). As long as journals serve as the basis for subject classification, journal indicators remain essential issues in bibliometric studies.

Table 13.8 Distribution of papers over performance classes by individual authors with at least 30 scientometric papers in 1998–2012. The χ^2 test is based on four CSS classes

Author ID	Papers	Class 1 (%)	Class 2 (%)	Class 3 (%)	Class 4 (%)	Class 3&4 (%)	χ^2	Type
309	65	32.3	33.8	10.8	23.1	33.8	104.04	IV
310	39	25.6	35.9	20.5	17.9	38.5	54.49	IV
442	80	46.3	32.5	13.8	7.5	21.3	20.67	IV
460	38	68.4	26.3	5.3	0.0	5.3	1.36	III
532	45	42.2	35.6	8.9	13.3	22.2	24.23	IV
585	88	53.4	29.5	12.5	4.5	17.0	9.95	IV
654	52	53.8	40.4	5.8	0.0	5.8	9.71	I
665	35	82.9	17.1	0.0	0.0	0.0	5.00	III
668	86	36.0	47.7	10.5	5.8	16.3	39.47	IV
702	48	47.9	37.5	6.3	8.3	14.6	11.85	IV
732	66	69.7	18.2	9.1	3.0	12.1	1.44	III
881	85	25.9	36.5	16.5	21.2	37.6	133.94	IV
1098	89	70.8	16.9	7.9	4.5	12.4	2.65	III
1109	34	44.1	41.2	11.8	2.9	14.7	9.07	IV
1538	41	68.3	29.3	2.4	0.0	2.4	2.89	III
1539	34	67.6	29.4	2.9	0.0	2.9	2.23	III
1826	45	33.3	42.2	17.8	6.7	24.5	25.98	IV
1962	38	34.2	47.4	13.2	5.3	18.4	19.30	IV
1964	39	35.9	46.2	12.8	5.1	17.9	17.86	IV
2058	30	33.3	56.7	10.0	0.0	10.0	21.27	IV
2249	36	47.2	25.0	13.9	13.9	27.8	20.05	IV
3594	43	37.2	44.2	16.3	2.3	18.6	20.59	IV
4870	30	50.0	36.7	10.0	3.3	13.3	4.38	III
Total	11514	67.6	23.1	6.4	2.9	9.3	0.00	III

Data sourced from Clarivate Analytics Web of Science Core Collection

In our first applications in 1987 and 1988, mean citation rates for chemistry journals were calculated for a 5-year publication period, with the same citation window, and then gauged against the CSS standards set by subfields of chemistry. The share of uncited papers was still used as an additional class 0 below the poorly cited articles. Thus three characteristic scores provided five classes, class 0 though class 4. We later refrained from this solution because of a lack of robustness. The share of uncited papers strongly depends on both the citation windows, i. e., it decreases as the window increases, and the subject matter. The reason for the instability is that this is the only class that is defined on a fixed criterion (i. e., uncitedness), while all other classes are based on variable thresholds that depend on time window and research topic. The basic idea at that time was to provide additional information about the journals' position in a discipline beyond the usual ranking exercises. *Schubert* et al. [13.22] already applied CSS to journals on the basis of two scores only and with the purpose of gauging national contribution to journal impact but not to compare class distributions by journals. Now we will use the CSS class distributions as an alternative to journal ranking.

This also demonstrates a new and completely different application context. Although this type of application does not allow any linear ranking, and the interpretation of CSS class distributions is not always straightforward, the added value of the information and the more detailed picture of citation impact that we obtain compensates for the greater complexity. The surplus of information results from two sources that have been addressed many times in the scientometric literature: first, the subject-specific peculiarities of citation impact, and second, the various shapes of citation distributions underlying journal impact. *Glänzel* and *Moed* [13.11] and *Glänzel* [13.12] have given examples of journals with almost identical mean citation rates but different distribution shapes. We have computed mean citation rates (denoted by MCR) and CSS class distribution for all journals in 2013 using a 3-year citation window. We have applied the fractionation process for subjects according to the procedure described in Sect. 13.3.2, since many journals have multiple subject assignments. Table 13.9 gives the mean citation rate values of five selected journal pairs with almost identical citation impact (MCR) each but distinctly different profile types according to their CSS class distribu-

Table 13.9 Examples of mean citation rates (MCR) of selected journal pairs with almost identical citation impact but distinctly different profile types in 2013 with 3-year citation window according to *Glänzel* and *Thijs* [13.37]

Journal title	Papers	MCR	Class 1 (%)	Class 2 (%)	Class 3 (%)	Class 4 (%)
Advanced Energy Materials	157	31.92	4.5	29.3	33.1	33.1
Molecular Biology And Evolution	238	31.71	42.0	42.0	13.0	2.9
American Journal of Respiratory and Critical Care Medicine	460	15.84	42.4	20.4	23.0	14.1
BMC Medicine	267	15.75	28.1	52.8	17.6	1.5
Scientific Reports	2455	11.16	61.1	31.7	5.8	1.4
Nutrition Reviews	83	11.16	21.7	38.6	27.7	12.0
Journal of Lightwave Technology	578	5.60	51.6	29.4	13.8	5.2
Clinica Chimica Acta	364	5.60	72.0	22.3	4.4	1.4
International Journal of Number Theory	120	0.90	74.2	20.8	4.2	0.8
Advances in Materials Science and Engineering	248	0.90	97.2	2.8	0.0	0.0

Data sourced from Clarivate Analytics Web of Science Core Collection

tions. We have chosen journals representing different *standards* to illustrate that this phenomenon might occur in all impact classes ranging from high to low standards.

Now we have a closer look at CSS profile types at the journal level. Also, for scientific journals, which correspond to the meso level in bibliometric practice, we can clearly distinguish all five types. The annual number of citable papers (articles, letters, notes, reviews) published in journals ranges from less than 10 to almost *industrial* quantities of several thousand in the major chemistry and physics journals, led by PLOS One with more than 30 000 papers in 2013. We have excluded the 5688 out of 12 386 journals with fewer than 50 papers each. Journals whose profiles did not deviate significantly from the reference standard in 2013 (70.0%–21.4%–6.2%–2.4%) were considered to be of type III. The distribution of journals over CSS profile types is thus as follows: 400 (type I), 3037 (type II), 1748 (type III), 1499 (type IV) and 14 (type V). Type V is thus extremely rare at this level too. Table 13.10 shows a sample of 30 journals across several subject fields per type in alphabetical order. While *Physical Review C* follows the reference standard almost perfectly, the deviation from this standard may be enormous as, e. g., *Accounting Review*, with quite homogeneous citation patterns, and *Prenatal Diagnosis*, with extremely polarized citation rates, clearly illustrate. We have not displayed the MCR values in order to avoid biased interpretation caused by mean citation rates. Nevertheless, type IV represents the most advantageous profile and type II the least favorable situation. Type III simply means general conformity to the subject-based standard. Figure 13.7 presents data on a selection of five triplets of the journals displayed in Table 13.10 just to visualize the variety of profiles at the journal level.

13.3.6 Application of Characteristic Scores and Scales to Citation-Impact Normalization

In an earlier paper, *Schubert* et al. [13.22] had observed that the difference $b_2 - b_1$ was a very close proxy for the standard deviation of the underlying distribution. This observation was based on a mere empirical finding, in particular, with the important property that the approximation improves with increasing sample size. However, the basic idea behind this formula is the fact that mean values of empirical Pareto-type distributions are approximately normally distributed, provided that the characteristic parameter α is greater than 2 [13.10]. Now, if we assume the very simple model in which the underlying distribution is approximately a Pareto-type distribution of Lomax type, then $\Delta\ b_k = b_k - b_{k-1} \approx ba^{k-1}$, and thus we have $b_2 - b_1 \approx ba$, where $a = \alpha/(\alpha - 1)$. In verbal terms, $b_2 - b_1$ is a linear function of b, where the slope depends only on α that is on the tail property of the underlying distribution. Yet the standard deviation of a Lomax distribution is $b\{a/(2-a)\}^{1/2}$. Both expressions are proportional to b and, if α is around 3, the coefficients amount to 1.5 and 1.73, respectively. The small deviation of the two values is less relevant, if all samples share the same α, as the most important property is the proportionality to b. On the basis of both the empirical evidence and the theoretical explanation, the following standardization of mean citation rates seems to be justified

$$u = \frac{x - b_1}{b_2 - b_1}\,, \tag{13.2}$$

where x represents the actual mean citation rate, and the u value is called the *unified citation score*. *Schubert* et al. [13.22] have introduced the unified citation score

Table 13.10 A sample of 30 journals across several subject fields per type in alphabetical order in 2013 with a 3-year citation window

Journal title	Papers	Class 1 (%)	Class 2 (%)	Class 3 (%)	Class 4 (%)	χ^2	Type
Accounting Review	170	43.5	49.4	5.9	1.2	80.66	I
American Journal of Occupational Therapy	59	52.5	45.8	1.7	0.0	22.34	I
Best Practice & Research: Clinical Gastroenterology	108	58.3	38.0	2.8	0.9	19.04	I
Biology of Blood and Marrow Transplantation	450	54.2	40.0	4.9	0.9	94.63	I
Ecotoxicology	1394	58.5	37.0	3.9	0.5	218.06	I
Electric Power Systems Research	1889	56.5	37.0	5.8	0.7	286.81	I
Acta Physica Sinica	2108	96.2	3.8	0.0	0.0	689.50	II
ECS Journal of Solid State Science and Technology	454	82.6	13.9	2.9	0.7	36.12	II
European Physical Journal B	528	90.3	8.5	0.9	0.2	106.22	II
General and Comparative Endocrinology	356	77.2	20.2	2.2	0.3	18.55	II
International Journal of Cardiology	2550	77.0	18.3	4.2	0.5	82.70	II
RSC Advances	3847	76.3	20.6	2.7	0.4	164.10	II
Applied Mathematics Letters	175	63.4	25.7	7.4	3.4	3.79	III
Circuits Systems and Signal Processing	178	77.0	19.7	2.8	0.6	7.30	III
Graefes Archive	408	69.4	23.0	6.4	1.2	3.00	III
Journal of Thoracic and Cardiovascular Surgery	580	66.9	23.3	8.1	1.7	6.42	III
Physical Review C	1068	70.4	21.4	5.7	2.4	0.41	III
Respiratory Medicine	254	69.3	23.6	5.9	1.2	2.27	III
Advanced Materials	785	10.2	30.6	29.2	30.1	3557.40	IV
American Journal of Epidemiology	371	41.8	34.0	18.1	6.2	176.07	IV
Analytical Chemistry	1477	33.6	34.3	18.6	13.6	1515.98	IV
Investigative Ophthalmology & Visual Science	1055	35.5	37.0	19.5	8.0	734.92	IV
Nanoscale	1546	29.8	42.8	20.3	7.1	1326.77	IV
Nutrients	148	34.5	43.2	12.2	10.1	104.64	IV
British Medical Journal	913	76.3	14.6	7.4	1.6	29.66	V
Discrete Mathematics	330	78.2	12.7	7.0	2.1	15.14	V
Eye	296	77.4	13.2	6.4	3.0	12.08	V
Journal of Analytical Toxicology	101	73.3	11.9	11.9	3.0	9.84	V
Prenatal Diagnosis	179	75.4	13.4	6.1	5.0	11.06	V
Journal of the Experimental Analysis of Behavior	50	78.0	6.0	10.0	6.0	9.80	V

Data sourced from Clarivate Analytics Web of Science Core Collection

in the context of the comparative analysis of national research performance. Obviously, the expectation of the unified citation score is 0 if the expected value of x is identical to that of the population, i. e., with b_1. If we prefer scores that take only nonnegative values, we can simply use the following random variable u^* instead of u. Then the same scaling effect is maintained, but the effect of uncitedness results in the value of zero as the lower bound of the normalized distribution,

$$u^* = \frac{x}{b_2 - b_1} \,. \tag{13.3}$$

Glänzel [13.20, 23] showed the robustness of this transformation for subfields and journal impact using examples of different publication years and citation windows, ranging from 3 to 21 years. Here we do not further deepen the issue of subject normalization using CSS, since scientometric research has come up with new, more sophisticated normalization solutions regarding a priori and a posteriori, i. e., cited- and citing-side normalization [13.24, 38–43]. Therefore, we mention the above property rather as one of the possible alternatives to a posteriori normalization and, most notably, as important evidence of the consistency of the Characteristic Scores and Scales method. However, with this issue, we arrive at the bottom of Fig. 13.2, which indicates the possible application of this kind of normalization in the context of research evaluation, and within that area, above all, at the lower aggregation levels.

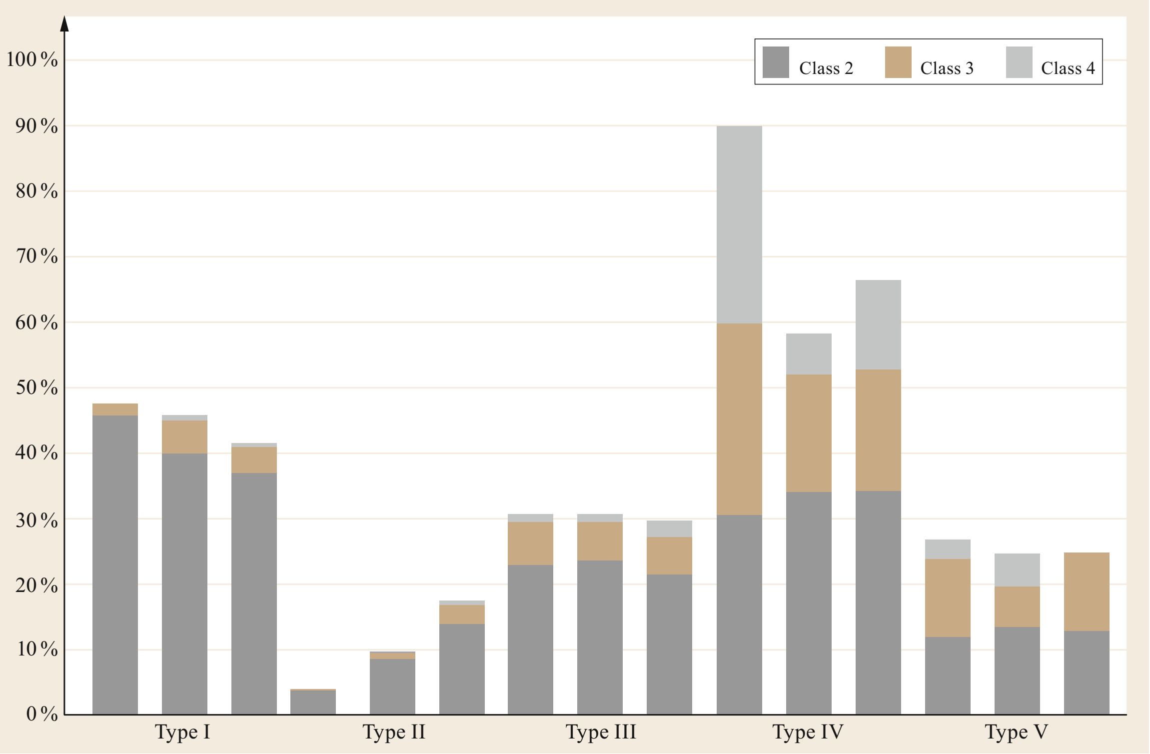


Fig. 13.7 Graphic presentation of a selection of five triplets of journals displayed in Table 13.10 (Data sourced from Clarivate Analytics Web of Science Core Collection)

13.4 Characteristic Scores and Scales in New Environments? Some Future Perspectives

In light of the foregoing, the question arises as to possible application of the CSS method in new, emerging environments. The so-called altmetrics, and more generally Scientometrics 2.0 [13.44, 45], are not only new challenges to bibliometricians; they also bring us new fields of application of methods previously devised and designed for their application contexts in the traditional "Scientometrics 1.x" [13.46]. In principle, the CSS method or its modifications or derivatives could be applied to any field that deals with similar phenomena of long-tailed Pareto-type distributions and statistics that might be biased by extreme values and outliers. The question as to how far the principles, rules and methods of Lotkaian informetrics [13.47, 48] also hold in their extension to altmetrics and social metrics is still to be answered. *Chi* and *Glänzel* [13.49] recently found that CSS could be applied to usage statistics provided by Clarivate Analytics (formerly Thomson Reuters) Web of Science Core Collection. According to *Pringle* [13.50, p. 1]:

> This count measures the level of interest in a specific item on the Web of Science platform. The count reflects the number of times the article has met a user's information needs as demonstrated by clicking links to a full-length article at the publisher's website (via direct link or OpenURL) or by saving the metadata for later use.

Unlike altmetrics indicators, which are designed rather to measure communication with an impact on the general public, usage counts focus again on communication among scholars [13.50]. In the above-mentioned pilot study by *Chi* and *Glänzel* [13.49], five disciplines from the life sciences, the natural sciences and the social sciences according to the Leuven–Budapest classification scheme and three countries were chosen to study whether usage and citations show similar patterns and to what extent these patterns might correlate at this level of aggregation. Again, the Characteristic Scores and Scales approach has shown the expected stability, with CSS class distributions very close to the usual 70%–21%–6.5%–2.5% rule.

Furthermore, the question arises of whether the observed properties of the Characteristic Scores and Scales method can be found in completely different contexts, for instance, in the context of degree distributions in network analysis or in graph models, or possibly in other fields of the sciences and socials sciences as well. Imaginable application fields outside informetrics could, for example, be economics, insurance mathematics or quantitative linguistics, i. e., fields where the (generalized) Waring distribution already plays an important role in modeling, identifying and processing a small quantity of outstanding and extreme observations, which in turn is contrasted by the otherwise overwhelming share of *low-profile* observations. We sincerely hope that the methods discussed in this chapter will help to tackle the pending and upcoming, yet unresolved, questions in both Scientometrics 2.0 and other contexts, wherever a power law, or even more generally a Paretian model, can be assumed to underlie the phenomena under study.

13.A Appendix

Table 13.11 List of country names and their ISO codes

Country	ISO code
Australia	AUS
Belgium	BEL
Brazil	BRA
Czech Republic	CZE
Denmark	DNK
France	FRA
Germany	DEU
India	IND
Israel	ISR
Italy	ITA
Japan	JPN
Netherlands	NLD
China PR	CHN
Poland	POL
Russia	RUS
Singapore	SGP
South Africa	ZAF
South Korea	KOR
Spain	ESP
Sweden	SWE
Switzerland	CHE
Taiwan	TWN
Turkey	TUR
UK	GBR
USA	USA

References

13.1 W. Glänzel, A. Schubert: Theoretical and empirical studies of the tail of scientometric distributions. In: *Informetrics 87/88*, ed. by L. Egghe, R. Rousseau (Elsevier, Amsterdam 1988) pp. 75–83

13.2 W. Glänzel: High-end performance or outlier? Evaluating the tail of scientometric distribution, Scientometrics **97**(1), 13–23 (2013)

13.3 W. Glänzel, K. Debackere, B. Thijs: Citation classes: A novel indicator base to classify scientific output. Paper presented at the OECD Blue Sky Forum, Ghent, 19–21 September 2016. Available via: https://www.oecd.org/sti/051%20-%20Blue%20Sky%20Biblio%20Submitted.pdf (2016)

13.4 W. Glänzel, A. Schubert: Characteristic scores and scales in assessing citation impact, J. Inf. Sci. **14**(2), 123–127 (1988)
13.5 A. Bookstein: Informetric distributions. 3. Ambiguity and randomness, J. Am. Soc. Inf. Sci. **48**(1), 2–10 (1997)
13.6 H.-P. Kriegel, P. Kröger, A. Zimek: Outlier detection techniques. Tutorial. In: *13th Pac.-Asia Conf. Knowl. Discov. Data Min (PAKDD), Bangkok* (2009)
13.7 L. Waltman, C. Calero-Medina, J. Kosten, E.C.M. Noyons, R.J.W. Tijssen, N.J. van Eck, T.N. van Leeuwen, A.F.J. van Raan, M.S. Visser, P. Wouters: The Leiden ranking 2011/2012: Data collection, indicators, and interpretation. In: *Proc. STI 2012*, ed. by E. Archambault, Y. Gingras, V. Lariviere (Science-Metrix, Montreal 2012) pp. 791–802
13.8 G. Matthys, E. Delafosse, A. Guillou, J. Beirlant: Estimating catastrophic quantile levels for heavy-tailed distributions, Insur. Math. Econ. **34**(3), 517–537 (2004)
13.9 W. Glänzel, K. Debackere: On the opportunities and limitations in using bibliometric indicators in a policy relevant context. In: *Bibliometric Analysis in Science and Research: Applications, Benefits and Limitations*, ed. by R. Ball (Publication Forschungszentrum Jülich, Jülich 2003)
13.10 W. Glänzel, H.F. Moed: Opinion paper: Thoughts and facts on bibliometric indicators, Scientometrics **96**(1), 381–394 (2013)
13.11 W. Glänzel, H.F. Moed: Journal impact measures in bibliometric research, Scientometrics **53**(2), 171–193 (2002)
13.12 W. Glänzel: The multi-dimensionality of journal impact, Scientometrics **78**(2), 355–374 (2009)
13.13 W. Glänzel, K. Debackere: On the "multi-dimensionality" of ranking and the role of bibliometrics in university assessment. In: *Ranking Universities*, ed. by C. Dehon, D. Jacobs, C. Vermandele (Editions de l'Université de Bruxelles, Brussels 2009) pp. 65–75
13.14 J. Beirlant, W. Glänzel, A. Carbonez, H. Leemans: Scoring research output using statistical quantile plotting, J. Informetrics **1**(3), 185–192 (2007)
13.15 L. Leydesdorff, L. Bornmann, R. Mutz, T. Opthof: Turning the tables on citation analysis one more time: Principles for comparing sets of documents, J. Am. Soc. Inf. Sci. Technol. **62**(7), 1370–1381 (2011)
13.16 L. Waltman, M. Schreiber: On the calculation of percentile-based bibliometric indicators, J. Am. Soc. Inf. Sci. Technol. **64**(2), 372–379 (2013)
13.17 M.A. Hamdan: On a characterization by conditional expectation, Technometrics **14**, 497–499 (1972)
13.18 S. Kotz, D.N. Shanbhag: Some new approaches to probability distributions, Adv. Appl. Probab. **12**, 903–921 (1980)
13.19 W. Glänzel, A. Telcs, A. Schubert: Characterization by truncated moments and its application to Pearson-type distributions, Z. Wahrscheinlichkeitstheor. Verwandte Geb. **66**, 173–183 (1984)
13.20 W. Glänzel: Characteristic scores and scales. A bibliometric analysis of subject characteristics based on long-term citation observation, J. Informetrics **1**(1), 92–102 (2007)
13.21 A. Schubert, W. Glänzel: A systematic analysis of Hirsch-type indices for journals, J. Informetrics **1**(3), 179–184 (2007)
13.22 A. Schubert, W. Glänzel, T. Braun: Scientometric datafiles. A comprehensive set of indicators on 2649 journals and 96 countries in all major fields and subfields 1981–1985, Scientometrics **16**(1–6), 3–478 (1989)
13.23 W. Glänzel: The application of characteristic scores and scales to the evaluation and ranking of scientific journals, J. Inf. Sci. **37**(1), 40–48 (2011)
13.24 W. Glänzel, B. Thijs, K. Debackere: The application of citation-based performance classes to the disciplinary and multidisciplinary assessment in national comparison and institutional research assessment, Scientometrics **101**(2), 939–952 (2014)
13.25 W. Glänzel, B. Thijs, P.S. Chi: The challenges to expand bibliometric studies from periodical literature to monographic literature with a new data source: The Book Citation Index, Scientometrics **109**(3), 2165–2179 (2016)
13.26 W. Glänzel, A. Schubert: A new classification scheme of science fields and subfields designed for scientometric evaluation purposes, Scientometrics **56**(3), 357–367 (2003)
13.27 B.C. Peritz: Are methodological papers more cited than theoretical or empirical ones? The case of sociology, Scientometrics **5**(4), 211–218 (1983)
13.28 W. Marx, L. Bornmann: On the causes of subject-specific citation rates in Web of Science, Scientometrics **102**(2), 1823–1827 (2015)
13.29 P. Albarrán, J. Ruiz-Castillo: References made and citations received by scientific articles, J. Am. Soc. Inf. Sci. Technol. **62**(1), 40–49 (2011)
13.30 I. Vincze: *Mathematical Statistics*, University Script, 4th edn. (Eötvös University, Budapest 1974), in Hungarian
13.31 W. Glänzel, A. Schubert, B. Thijs, K. Debackere: Subfield-specific normalized relative indicators and a new generation of relational charts: Methodological foundations illustrated on the assessment of institutional research performance, Scientometrics **78**(1), 165–188 (2009)
13.32 B. Thijs, K. Debackere, W. Glänzel: Improved author profiling through the use of citation classes, Scientometrics **111**(2), 829–839 (2017)
13.33 P. Wouters, W. Glänzel, J. Gläser, I. Rafols: The dilemmas of performance indicators of individual researchers – An urgent debate in bibliometrics, ISSI Newsletter **9**(3), 48–53 (2013)
13.34 W. Glänzel, P. Zhou: Publication activity, citation impact and bi-directional links between publications and patents in biotechnology, Scientometrics **86**(2), 505–525 (2011)
13.35 W. Glänzel: On co-authorship at the individual level, Transinformação **26**(3), 229–238 (2014)
13.36 A. Schubert, W. Glänzel, T. Braun: Subject field characteristic citation scores and scales for assessing research performance, Scientometrics **12**(5/6), 267–291 (1987)

13.37 W. Glänzel, B. Thijs: The granularity of disciplinary structures for benchmarking citation impact. The case of CSS profiles. In: *Proc. ISSI 2015 – The 15th Int. Conf. Scientometr. Informetr. (Wuhan)* (2017) pp. 1190–1200
13.38 H.F. Moed: Measuring contextual citation impact of scientific journals, J. Informetr. **4**(3), 265–277 (2010)
13.39 M. Zitt, H. Small: Modifying the journal impact factor by fractional citation weighting: The audience factor, J. Am. Soc. Inf. Sci. Technol. **59**(11), 1856–1860 (2008)
13.40 M. Zitt: Citing-side normalization of journal impact: A robust variant of the audience factor, J. Informetrics **4**(3), 392–406 (2010)
13.41 M. Zitt: Behind citing-side normalization of citations: Some properties of the journal impact factor, Scientometrics **89**(1), 329–344 (2011)
13.42 L. Waltman, N.J. van Eck: A systematic empirical comparison of different approaches for normalizing citation impact indicators, J. Informetrics **7**(4), 833–849 (2013)
13.43 L. Waltman, N.J. van Eck: Source normalized indicators of citation impact: An overview of different approaches and an empirical comparison, Scientometrics **96**(3), 699–716 (2013)
13.44 J. Priem, B.H. Hemminger: Scientometrics 2.0: New metrics of scholarly impact on the social Web, First Monday (2010), https://doi.org/10.5210/fm.v15i7.2874
13.45 J. Priem: Altmetrics. In: *Beyond Bibliometrics: Harnessing multidimensional indicators of scholarly impact*, ed. by B. Cronin, C.R. Sugimoto (MIT Press, Cambridge 2015)
13.46 W. Glänzel, P.S. Chi: Scientometrics 2.0 – And beyond? Background, promises, challenges and limitations, ISSI Newsletter **12**(3), 33–36 (2016)
13.47 L. Egghe: *Power Laws in the Information Production Process: Lotkaian Informetrics* (Emerald Group Limited, Bingley 2005)
13.48 L. Egghe: Lotkaian informetrics and applications to social networks, Bull. Belg. Math. Soc. – Simon Stevin **16**(4), 689–703 (2009)
13.49 P.-S. Chi, W. Glänzel: An empirical investigation of the associations among usage, scientific collaboration and citation impact, Scientometrics **112**(1), 403–412 (2017)
13.50 J. Pringle: *The Web of Science Usage Count* (Thomson Reuters, 2015) available at: http://thomsonreuters.com/content/dam/openweb/documents/pdf/scholarly-scientific-research/white-paper/web-of-science-usage-countreport.pdf

Wolfgang Glänzel

ECOOM and Faculty of Economics and Business
KU Leuven
Leuven, Belgium
wolfgang.glanzel@kuleuven.be

Wolfgang Glänzel is Director of the Centre for R&D Monitoring (ECOOM) and Full Professor at KU Leuven. He is also affiliated with the Department. Science Policy and Scientometrics at the Library of the Hungarian Academy of Sciences in Budapest (Hungary). Wolfgang Glänzel studied mathematics at Eötvös Lorand University (ELTE) in Budapest. He holds a Doctorate in Mathematics from ELTE, obtained in 1984, as well as a PhD in the Social Sciences obtained from Leiden University (The Netherlands) in 1997.
His research activities comprise probability theory and mathematical statistics, quantitative science studies, and research policy. Wolfgang Glänzel has published numerous journal articles, proceedings, and book chapters, and has co-authored/co-edited several books. He is Secretary-Treasurer of the *International Society for Scientometrics and Informetrics* (ISSI), *Editor-in-Chief of the international journal Scientometrics, and Academic Editor of the multidisciplinary OA journal PLoS ONE.*

Bart Thijs

ECOOM
KU Leuven
Leuven, Belgium
bart.thijs@kuleuven.be

Bart Thijs is a Senior Researcher in Bibliometrics at KU Leuven – ECOOM. His research focusses on the identification of global and local network structures within scientific communication and on the development and application of citation-based indicators. He combines a background from Mathematical Psychology (Master's degree from KU Leuven) and Social Sciences (PhD from Leiden University) with professional experience as a statistical consultant in his work.

Koenraad Debackere

ECOOM and Faculty of Economics and Business
KU Leuven
Leuven, Belgium
koenraad.debackere@kuleuven.be

Koenraad Debackere is a Professor of Innovation Management and Policy at KU Leuven. He is Promotor-Coordinator of ECOOM. His research has focused on the development and use of indicators measuring the effects of science and innovation policy, the role of entrepreneurial universities in economic development and the management and growth of knowledge-intensive companies.

14. An Overview of Author-Level Indicators of Research Performance

Lorna Wildgaard

The purpose of this chapter is to present a critical overview of author-level indicators of research production (ALIRP), discuss their appropriate application and provide a tool to support the informed use of ALIRP. A brief history of the development of ALIRP begins with a chronological discussion of the major trends in indicator development, which documents the quick adaptation of ALIRP in evaluation practice, and consequently sets the argument for the need to monitor and evaluate present-day indicator production, which is the major theme of this chapter. The characteristics and common mathematical properties of ALIRP are used to highlight the challenges we face in applying appropriate ALIRP in evaluation. The construction and validity of 69 ALIRP are analyzed, and the results presented in table form for easy reference. These tables are also available as interactive tables provided as e-material to this chapter. This analysis, combined with the deconstruction of indicators in the chapter sections, argues that ALIRP are mathematical models, and the numerical values they produce should never be confused with the reality they are trying to model in evaluation practice.

14.1 A Brief Introduction to Author-Level Indicators

The author-level indicator market is a crowded marketplace. Since the introduction of the *h*-index in 2005 [14.1], the number of indicators that measure some aspect of the individual researcher's published output continues to grow exponentially. However, few have attempted to systematically create an overview of the indicator products available in this market [14.2–6]. The proliferation of indicators is well intentioned but not always well informed. Indicators can be ill-applied in practice, leading to distrust of bibliometrics as an evaluation method. The purpose of this chapter is to present a critical overview of author-level indicators of research performance (ALIRP), with particular emphasis on newer developments toward disciplinary- and seniority-specific indicators, and the practical challenges we face in the context of their application. The chapter does not consider alternative quantitative author-level indicators such as Altmetrics, econometrics or network metrics, nor does it address qualitative evaluation indicators of individual performance. This approach can perhaps be regarded as a narrow view that does not fully reflect the plurality of the indicator research. It is nonetheless a realistic view addressing the

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-02511-3_14) contains supplementary material, which is available to authorized users.

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_14

practical application of indicators, and highlights the pressing matter of indicator validation. Accordingly, it is important to start off this chapter by defining ALIRP and what they are considered to measure—or, importantly, not measure—and refer briefly to the ongoing discussion in the bibliometric and research evaluation community.

According to *van Raan* [14.7]

> An indicator is a measure that explicitly addresses some assumption. [...]. Indicators cannot exist without a specific goal in mind, they have to address specific questions, and thus they have to be created to gauge important *forces*. [...]. Indicators must be problem driven, otherwise they are useless.

The aim of indicators is to provide a better understanding of how science works, how it can be measured and the extent to which indicators can inform evaluation practice. Hence, bibliometric indicators are designed to answer specific questions at specific aggregation levels. Aggregation can occur at the level of the individual author, group of authors, institutions, journal, field or country, among others. As each level of aggregation has its own inherent properties, it is possible to distinguish indicators designed for these successive levels of aggregation from one another. Author-level indicators of research performance (ALIRP) are a set of quantitative indicators developed to capture aspects of quality, impact and prestige at the aggregation level of the individual researcher. The desired properties of ALIRP are suggested by *Todeschini* and *Baccini* [14.8, pp. 24–25] as follows:

1. Univocal mathematical definition
2. Rank top and young researchers in a balanced way
3. Sensitive to number and distribution of citations and papers
4. Robust to small variations in citations and papers
5. Ability to rank researchers (low degeneracy)
6. Preserve sensitivity also for top researchers
7. Easily computable from the available data.

At higher levels of aggregation, the consistency properties are different [14.9]. *Waltman* et al. draw our attention to the fact that the consistency properties of indicators applied at different levels of aggregation are not merely of interest theoretically, but also have significant practical relevance [14.9]. The practical relevance of ALIRP is an extremely important discussion. Not only are ALIRP used to summarize a researcher's set of papers and citations and to assess their contribution to a specific research community; the resulting values are also used as performance benchmarks within and across disciplines and seniority levels in research evaluation, hiring, promotion and funding decisions. The basic data used to compute the indicators are counts of the number of citations and publications, and they are value-laden data. What is defined as a publication or set of papers differs from research area to research area [14.10, 11]. Different types of publications represent different research methods, which in practice render comparisons of counts among different publications unfeasible [14.12]. Some consider publications as simply a discrete set of objects that can be counted in aggregate [14.13], while others maintain that publications must be limited to output indexed in citation databases for indicators to make sense [14.14]. Similarly, citations are open to interpretation—from recognition of intellectual debt [14.15], indications of authority [14.16], or markers of scientific communication [14.17] or cognitive influence [14.18], to simply "a quantitative and computer manipulable measure of *something or other*" [14.19]. Further, citations do not necessarily measure the extent to which important scientific output is communicated and used: as authors, we fail to cite, we cite to criticize, and we fail to acknowledge ideas/intellectual influence [14.20, p. 127]. Both publications and citations have different values with regard to paradigmatic and social norms [14.21, 22], but when publication and citation counts are combined in bibliometric indicators, they produce numeric values that we *assume* tell us *something* about the researcher's past performance, *partially* inform us of scientific progress and/or produce a statement of *impact*. There is yet no consensus as to what these indicators are in fact measuring and, consequently, what we can claim. There is agreement, however, that in the realm of evaluation, publications are just *one* measurable aspect of scientific productivity and, together with citations, claim a role as intermediary devices in assessments of *aspects* of scientific quality. So, even before we discuss the construction and mathematical foundations of ALIRP, we have encountered the core problem that has vexed the bibliometric community for years. How do we define a publication, a citation or even an author, and operationalize these concepts as countable variables in performance indicators? Clear definition does not just ensure robust and valid operationalization of variables included in ALIRP. Clear definition ensures responsible interpretation of the numbers these indicators produce and enables informed statements about the relations between an individual's productivity, impact, collaboration and prestige, among other qualities.

The methods we use to count and combine publications and citations are not neutral techniques. Even the databases from which we harvest bibliometric data originate from Northern European scientific communities, and thus primarily reflect the practice characteristics of the natural sciences in these regions. In the construction of author-level indicators, specifically as instruments to measure performance, we must be aware of the strengths and limitations of applying indicators as a positivistic approach to research evaluation. The strength of the positivistic approach is that it is intended to prevent confusion, rejecting as meaningless all concepts that cannot be verified by experiment and logical analysis. In indicator development, this means that indicators use what can be counted to verify clearly defined *partial* aspects of an individual's performance. What is countable is regarded as empirical fact. The first main weakness of the positivistic approach is that the sociological and theoretical variables of researcher output and activity that do not produce quantifiable data remain invisible. In strict adhesion to the positivistic ethos, these should be considered meaningless at any rate. Common sense tells us this is not so. Philosophical and sociological discourse is essential in legitimate performance evaluation, but the research on indicator development presented in this chapter does illustrate a clear demarcation between indicator development and indicator application in practical research assessments. Secondly, confusion does not appear to be prevented. There is great tension between detail and accuracy: an inverse relationship perhaps between the precision of the bibliometric data and the substantiation of the indicator [14.23]. The bibliometric community appears divided between those that develop indicators and those that apply indicators in evaluation practice. This division is not necessarily a bad thing; it sparks the creativity of the inventors of indicators and encourages critical use in the application context, as we will explore further in the next sections.

14.2 Brief Review: Trends in Indicator Development

In the 1960s and 1970s, Derek J. de Solla Price published a number of books and articles that, together with the development of Garfield's Science Citation Index, laid the foundation for the emerging field of quantitative science studies as a research program [14.24]. Yet it was only after the journal *Scientometrics* was founded by Tibor Braun in 1978 that publications of studies on how to measure science were truly liberated from traditional information, technology and computer science journals. That same year saw an introduction of best practices in exploring broad, new perspectives and possibilities for science indicators, published in a book by *Elkana* et al. [14.20], featuring the writing of leading historians, philosophers of science and social scientists. The authors emphasize the importance of practice standards, particularly with respect to taking data seriously in a historical context in order to produce responsible indicators based on science, rather than on group politics under the guise of science. These are the same principles that would be reiterated nearly 30 years later in the Leiden Manifesto [14.25] and in the Metric Tide [14.26]. In the context of ALIRP, *Elkana* et al. [14.20, p. 132] discuss the duality of applying mechanistic indicators in what they argue should be primarily a humanistic evaluation of individual researchers. Researchers are viewed as complex systems of human activity and relationships that together result in scientific products and services. Generalizations, therefore, based on the history and sociology of science and theoretical approaches to measuring science break down when tested on individual cases and examined in sufficient detail. Humanistic explanations are far more powerful than those furnished by quantitative indicators at the individual level, and certainly human activity should be explained humanistically rather than interpreted through mechanical, numerical concepts. Measuring science mechanistically, counting production for example, directs the evaluation to the authority of the techno-structure rather than the authority of the evaluand (the person under assessment). The evaluation is focused on the network of managers, policymakers and administrators who control procedural decisions, strategies and economies both within and beyond the realm of the individual. It is distanced from the interests of the researcher and their goals, potential, consumers, or the fusion of their scientific production and wisdom. In a mechanistic approach, indicators may only be of value in monitoring and planning science *if* they are used with an awareness that they are but one of many inputs in an evaluation program. They alone do not reflect the vigor or potential of the researcher.

Nevertheless, the continued desire to use algebraic models in a mechanistic approach to quantify and compare the production, vigor, prestige, outreach and potential of researchers and to counterbalance the role of peer review in evaluation has fueled the creation of an enormous number of indicators. Already in 1978, a call for caution was sent out by *Elkana* et al. [14.20], along with a plea for prioritizing responsible and creative science indicators to inform science policy. Figure 14.1 illustrates the growth in papers concerning the development of ALIRP, (reference search in Web of Science, Scopus

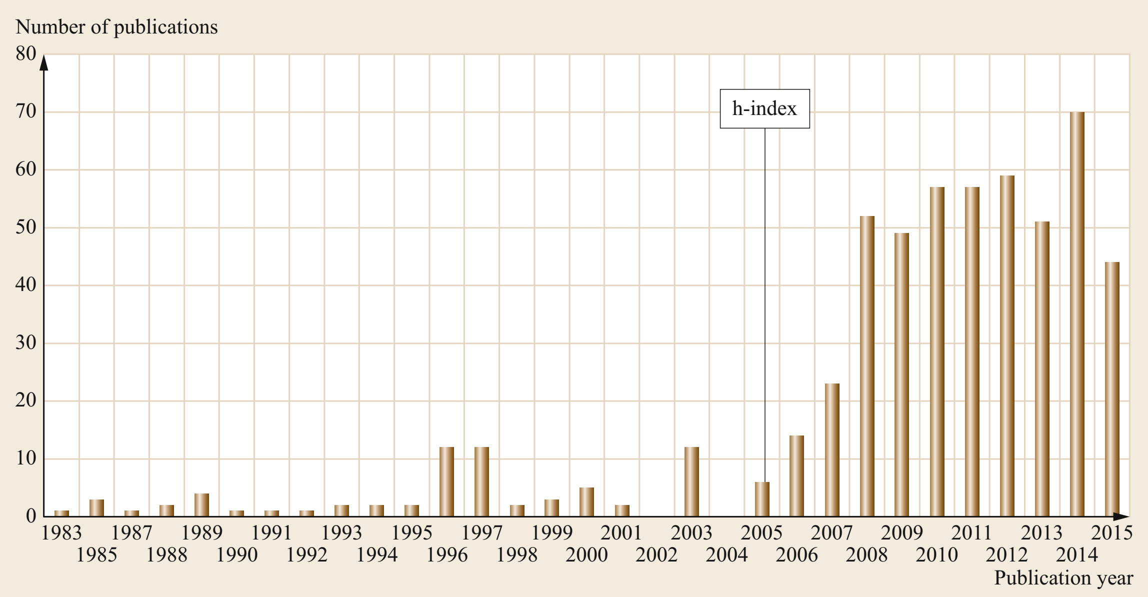


Fig. 14.1 Development of ALIRP, 1980–2015

and Google Scholar, 1980–2015, undertaken December 2016).

The more or less immediate response of the bibliometric community in the 1980s was the development of ALIRP that concentrated on the correlation between bibliometrics and peer judgment [14.27–30], and on how to account for the differences between researchers in different disciplines in scholar rankings and performance evaluation [14.31–33]. In the mid-1990s, bibliometricians argued that although citation indices provided a logical starting point for measuring production and citation effect, the limitations of these sources could leave a void in the coverage of citations to an author's work, a void that must be explicitly recognized in evaluation. Consequently, a heated debate ensued regarding how to improve citation indexing to ensure the future role of the indices in responsible evaluation of individual researchers [14.34–36], leading to a number of exemplary investigative studies, though still primarily in the hard sciences [14.37–40]. Some fundamental questions were also being addressed—importantly, the performance of ALIRP in science and technology in order to analyze the dynamics between patents, publications and the movement of researchers between research institutions [14.41, 42], as well as the role of citations in communication theory, building upon earlier work [14.43–45].

The second half of the 1990s saw the blossoming of a discussion on author-level bibliometrics as a paradigm [14.45]; specifically, during this period, the groundwork was laid for the standardization of bibliometric terminology and indicators [14.46–49], and the operational definitions of authorship and collaboration in indicator development were heavily debated [14.50–53]. The potential for indicators to capture collaboration from a research evaluation and policy perspective is a recurrent theme to the present day [14.54–57], specifically the extent to which inter-institutional and international collaboration is beneficial for an individual's research production and career trajectory. Ultimately, the 1990s ended with its own call for a unified indicator theory [14.58–60] to ensure that the hunger for metrics from science administrators and policymakers was satisfied with robust and valid metrics. However, the call was not heeded. Instead, there was a shift away from theoretical discussions and the operationalization of citations in the development of ALIRP, towards the application of ALIRP by governmental agencies to measure innovation and demonstrate the value of productivity in management and funding policy strategies [14.61, 62]. Accordingly, the validation of indicators as appropriate measures of individual researcher performance became a major concern within the bibliometric community [14.63–66], who, in response to the increasing institutionalization of indicators and their somewhat unconstrained application in research evaluation, argued for the correct contextual application of indicators and the need for data quality assessment methods [14.10, 67–70]. Consequently, the bibliometric literature became increasingly concerned with the potential and limitations of alternative sources of bibliometric data as a supplement to or replacement for traditional citation indices at the individual level [14.71, 72]. Of particular interest was the potential of the recently established Google Scholar database as a citation index [14.73–75], as the influence of the scope of the citation index on author-level indicator values is not inconsequential in rankings [14.76–80].

Throughout the 1980s to mid-2000s, the development of ALIRP was strongly considered by researchers as *blasphemous* and by bibliometricians as a practice that should be avoided, because such indicators would be inadequate for evaluation and would lead to erroneous conclusions [14.81]. The cardinal rule of indicator development was that "the biases and deficiencies of individual citers are repaired to a tolerable degree by the combined activity of the many" [14.82], that is, small sets of bibliometric data, those representing individual researchers, would lead to statistical problems that would distort indicator values [14.83]. However, a fundamental change in the basic concepts and experimental practices in indicator development was heralded with the publication in 2005 of the h-index [14.1], a type of indicator never before seen, developed by a different kind person from outside the field of bibliometrics! A Hawkinesque indicator, a single unifying equation that would explain everything, able to balance the quantity of an individual's publications with impact, quality, prestige and rank among peers, the h-index quickly captured the attention of the scientific world, policymakers and the media, gaining legitimacy and acceptance as a useful measure from leading scientific journals [14.84]. Bibliometricians were quick to criticize the h-index for oversimplifying the conundrums inherent in the development and application of ALIRP, encouraging users to substitute mathematical rules for judgment in evaluation [14.3, 85]. But they were just as quick to cast off their previous reluctance to develop ALIRP, producing a barrage of new ALIRP, each claiming to be more robust, valid and sophisticated than the last [14.2, 86–88]. However, with no advisory boards, common standards or contextual assessments, the indicators were largely incomparable, laboratory experiments. They appeared theoretically unfounded, the methodological and operational origins often concealed from the users. Validity tests, if any, were conducted only on small choice sets of bibliometric data [14.81, 89].

The rapid adoption of ALIRP in research policy continues to steer the direction of indicator development to the present day [14.90–92]. First, the operationalization and institutionalization of ALIRP in hiring, reappointment, tenure and funding decisions [14.90, 93–97] has led to experiments with indicators that can objectively account for gender [14.98–100], seniority [14.101, 102] or disciplinary bias [14.103–106]. Second, the added value that ALIRP bring to evaluations has been investigated, especially when they are used as a supplement to traditional input–output investment indicators [14.107–110] in the evaluation system, along with their influence in the processes of researcher/departmental development. These advances into the political arena have forced the bibliometric community to readdress guidelines for meaningful evaluation at the individual level—a need first called for but fundamentally ignored in the 1990s [14.45–47, 49, 111]—and to communicate guidelines for the practical implementation of indicators for users outside the core bibliometric community [14.112–117]. Not surprisingly, the increased interest in the use of ALIRP in policy and evaluation has led to an increased demand for assessment of the validity and reliability of indicators at the individual level [14.90, 118–122]. As a result, a number of studies have investigated the psychological effects that bibliometric evaluations can have on the researcher, and reactive changes in publication behavior [14.123–127]. A process-oriented rather than diagnostic approach to accommodating change is seen in the current development of indicators, such as the benefits of publishing alone or in groups in evaluation of productivity, how to compare *mass* production to quality production, strategies in the author-byline hierarchy, and the challenges of name disambiguation in the attribution of credit [14.128–132]. Validation of indicators appears to have defined a shift in recent bibliometric literature, which seems to have turned increasingly introspective, examining whether the *appropriate* methodology is being used by bibliometricians to explain and predict trends in bibliometric analyses [14.133, 134]. In earlier years, this topic was limited to peripheral discussions [14.135–139]. Drawing on knowledge from the field of statistics, *Schneider* [14.133, 134, 140] and others [14.141–146] challenged the overreliance of the bibliometric community on sample statistics and false precision to argue the strength of the findings in bibliometric experiments. Bibliometrics is not a pure science, and at the individual level, indicator values cannot be detached from the person under evaluation. The shortcomings of the mechanistic approach and the strength of the humanistic approach suggested in 1978 by *Elkana* et al. are being revisited. ALIRP cannot be dissociated from the scientific and cultural materials the individual produces, as they are produced in a social system through complex relationships for the sake of science, not for the sake of a statistical method of measurement [14.147]. Determining which concepts are operationalized in indicators and how they are operationalized, where the data come from, what is missing and how sociological concepts can be interpreted without a theoretical framework constitutes a real problem [14.148–150]—a more important problem than sophisticated statistical calculations.

And so the literature has gone full circle, back to 1978 and the call for practice guidelines and fidelity to historical data [14.20]. The growth and interest in ALIRP has evolved at an astonishing rate, with a recent focus on developing responsible metrics and good practice guidelines for consumers of ALIRP. Concerns with practical implementation has perhaps taken attention away from the equally important need for theoretical and methodological frameworks that support the development of robust ALIRP and reduce the production of ad hoc ALIRP. Such frameworks could have a huge influence on how indicators are constructed and how indicators designed specifically for analysis at the individual researcher level are legitimized.

14.3 General Characteristics of Author-Level Indicators

In the previous section, prominent themes driving the development of ALIRP were discussed chronologically. This brief review led us to the current state of indicator development, where there is strong consumer demand from research administration and policymakers for quantitative measures to inform decisions. To support decision-making in science policy scenarios, ALIRP could be considered as *customer-value* models, which are data-driven representations of the worth, in mathematical terms, of the researcher under evaluation. To learn more about the robustness and validity of ALIRP in informing practice and decisions, it is crucial to have a shared understanding of how these customer-value models are built and the characteristics that define them.

14.3.1 Interdisciplinary Collaboration in the Development of ALIRP

Developers from outside the bibliometric research community are highly visible in the production of ALIRP. They use their disciplinary expertise to suggest indicators specifically designed to fit the characteristics of their field to inform responsible evaluation. In Fig. 14.2, which is adapted from *Wildgaard* [14.89, Ch. 6], the disciplinary partnerships among developers of 57 ALIRP are mapped. These developers represent 125 research specialties, which were determined using the authors' own keywords from their biographies, their CVs or institutional webpages. For simplification, these specialties are collapsed into 17 broad fields of study, represented by the circular nodes on the sociogram. Counts are not fractionalized, meaning that a developer can belong to more than one research area. The figure illustrates the extent to which ALIRP are produced through collaborative design, i. e., teams of developers with varying skill sets and expertise from different research areas, who work together on the design of an indicator to capture aspects unique to a field, specialty or academic profile in the indicator model. The *IQP* indicator, for example, is grounded in the fields of psychology, economics and leadership, whereas the *b*-index is developed in the field of chemistry. In practice, it may be wise to be mindful of the worldview of the developer, as this could determine field-specific goals of the indicator and the types of publications, authors and sources, for example, that are operationalized in the indicator model.

14.3.2 The Immaturity of Indicators

The greatest shared characteristic of ALIRP is their immaturity, both those presented in the bibliometric literature and those published in other disciplinary journals, blogs or publication channels. The majority of ALIRP remain tentative proposals, untouched from the day they were put forward, not tested in follow-up empirical studies. This has left us with a large set of indicators that remain theoretical. Theoretical indicators are important, as they help us gain greater knowledge of paradigmatic mathematical laws and properties of bibliometric distributions, e. g., hw, *dynamic h* and *tapered h* (hT), that in turn could eventually inform the development of ALIRP. Yet practical application of these theoretical (or new or unestablished) indicators in empirical studies is vastly neglected. We need to learn more about the suitability of indicators in different application contexts and the ability of the indicator, in combination with other evaluation measures, to capture the heterogeneity of scholarship. Empirical follow-up research has been published on the extent to which some indicators build on, overlap or supplement one another, primarily with regard to the robustness of the *h*-index [14.80, 88, 91, 109, 151–154] and suggestions for multiple design variants of *h* for specific disciplines and seniority levels [14.3, 86, 88, 101, 135, 153, 155]. However, this is far from sufficient to establish a practice of validation and verification of the many indicators we have at our fingertips. Further, these variants themselves represent suggested improvements that require verifica-

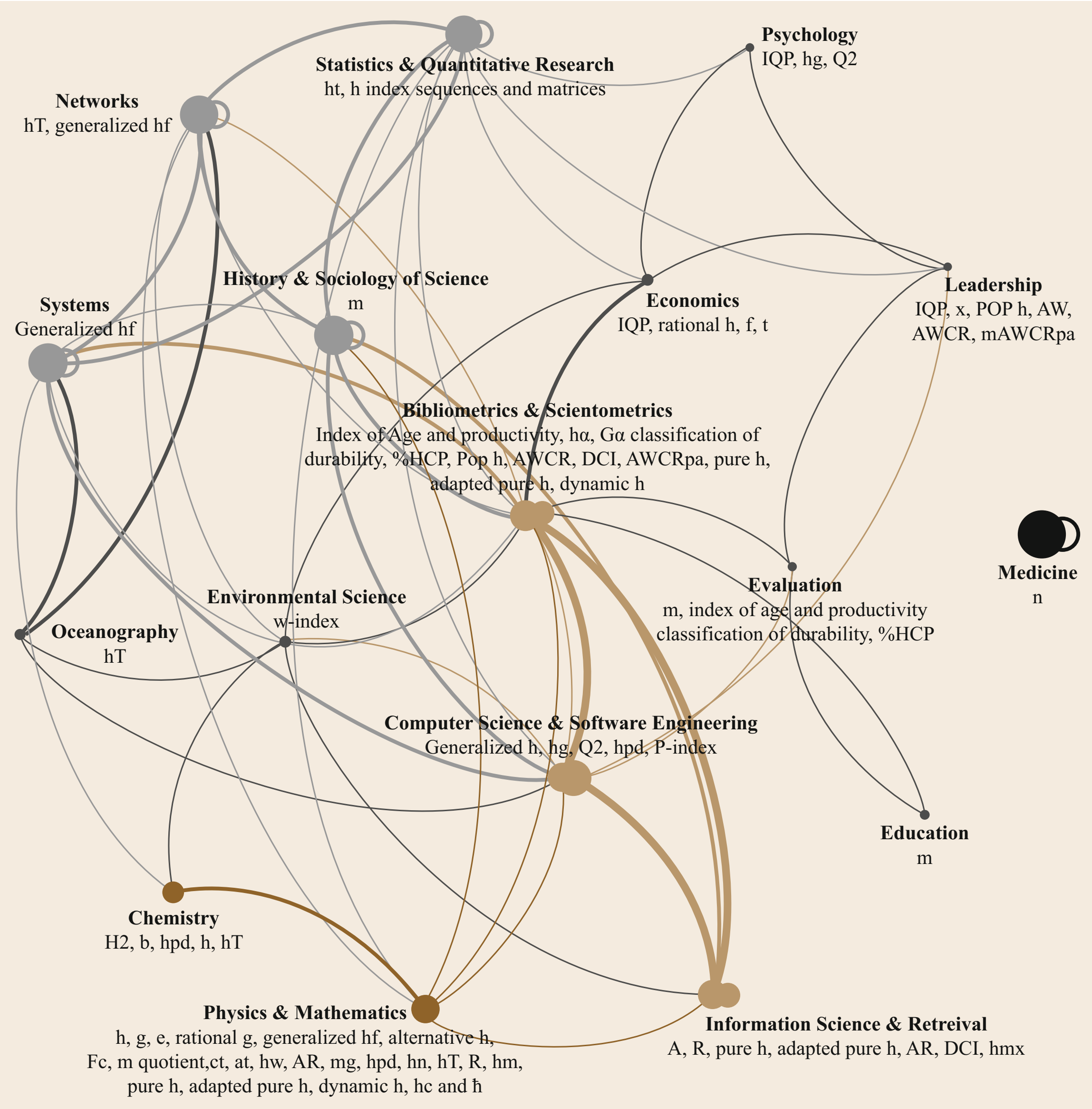


Fig. 14.2 Map of interdisciplinary collaboration between developers of 57 ALIRP

tion and validation, which means that caution should be exercised if they are applied in their present state in *real-life* evaluations.

14.3.3 Conceptual Operationalization and Model Validity

Methodological clarity is a basic requirement for the construction of all mathematical models. As models, ALIRP are representations of *something*, an analogy to help us visualize something such as an effect or impact, something that cannot be directly observed. As such, they are "a system of postulates, data and inferences presented as a mathematical description of an entity or state of affairs" [14.156]. Accordingly, they are cognitive activities that use mathematical concepts and language to describe how something behaves in real life. Therefore, it is important to validate them, to know how and when to use them, and to know when their use is limited. This entails 1) identifying how the developer(s) of the indicator conceptualized, defined and operationalized the variables in the indicator model (most commonly author, publication and citation, but there are others including time, seniority and gender), and 2) ascertaining how these variables are combined through different arithmetic functions to measure a specific aspect of research performance. Homogeneity and consistency, as discussed by *Gingras* [14.157], are central to ALIRP, in that every indicator we use must be dimensionally homogeneous and dimensionally consistent. The following is an example of the common-sense validation technique devised by *Gingras* [14.157]. The h-index is defined by *Hirsch* as a measure of the importance, significance and broad impact of an individual's overall research output, and is a composite measure of productivity (number of published articles) and citation effect (number of citations) [14.1]. This means that every term in the h-index should describe a relation to these two concepts, and should operationalize the total dimensions of importance and significance. Yet the design of the indicator is inconsistent with our understanding of the relationship between the quantity of papers and citations and the concepts of importance and significance. Also, the h-index claims to be an index designed to *quantify* a researcher's scientific *output*, and as such we expect an increase in h equal to the increase in the importance/significance of a researcher's work, and likewise a decrease if there is a decrease in the importance/significance of the researcher's work. Many studies have shown flaws in the inertia of h, situations where the lower h-index hides the better-quality researcher.

There is no doubt that mathematical models have been successfully developed and applied in scientometrics at higher levels of aggregation, providing insight, for example, into disciplinary production and the direction of research at a national level. Doubt remains, however, as to whether ALIRP can model events successfully at the individual researcher aggregation level, because the situation of the individual is incredibly complex or, as exemplified with the h index, because the models are intractable. For ALIRP, it is essential to be able to defend the relations among four basic facets of the indicator model. These are:

1. The aim of the indicator, i. e., the aspect of performance the indicator is designed to evaluate
2. The context of the person under evaluation (typically operationalized as an author)
3. The data (publications and citations)
4. The calculation methods (Fig. 14.3).

Figure 14.3 (first published in [14.89, Ch. 6]) illustrates the different approaches adopted by 51 ALIRP in operationalizing the variables of author, publication and citation, and the concept that the 51 indicators claim to measure. Full insight into the data is given in the electronic supplementary material that is provided with the online version of this chapter. Developers argue the adequacy of the indicator through the definitions and operationalization of author, publication and citation. Naturally, the definition and operationalization of these core variables vary, as indicator developers are attempting to capture different aspects of research performance in a very heterogeneous landscape. Similarly, the ALIRP could still measure what it is intended to measure even though the definitions of these variables are poorly articulated. It would be a mistake to conclude that an indicator that cannot isolate these variables of measurement has a certain level of defect. Poor articulation does not mean that the indicator should be dismissed as irrelevant or misleading. Rather, it highlights the need for developers and consumers of ALIRP to learn to evaluate the description, objectives and components of indicators, not just the indicator's strengths and weaknesses in evaluation contexts. The author concept informs us regarding the individual for whom the indicator model is designed (Fig. 14.3), be it an award winner, a scientist, someone who has published or someone who has both published and been cited. The concept of citation informs us as to how the developers interpret citations, and varies from indicator to indicator, whether as part of a reward system, recognition of use over time, quality, popularity or transference of ideas. The concept of publication informs us as to the specific type of output counted in the indicator, be it papers that have been cited, papers indexed in specific citation indices or papers from other resources. Some ALIRP are designed to include all of a researcher's publication output, while others advocate the practical usefulness of restricting papers to those in a citation index such as the Web of Science (WoS). Papers in WoS are argued to have an implicit aspect of quality that can be utilized in ALIRP to say something about some aspect of quality of the researcher's work, and enable comparisons of output among similar researchers. To be included in the citation index, papers have passed peer review and are published in mainstream disciplinary journals that have a certain level of citation, and further, the papers are represented by a bibliographical record that makes the paper searchable and verifiable. Likewise, the number of works citing a paper are also registered and details of these are indexed as well, enabling quantitative studies of scientific communication (for additional detail, see *Moed* [14.18, pp. 35–50]).

Together, the concepts of author, citation and publication are operationalized to measure something—the currency, the excellence, the growth, the quality and quantity of an individual's work—or to produce rank-

		Publication					Citation											Measure											
		Paper	Papers in WoS	Papers in other index	Object with citation	Expression	Effect	Hirsch definition	Influence	Impact	Performance	Popularity	Quality	Reward	Transfer of ideas	Use over time	No definition	Career	Comparison	Currency	Distribution	Durability	Effect	Excellence	Growth	Independence	Quality	Quality & Quantity	Rank
Author	Author	Adapted pureH	fc; ħ;Q2; b;hT	hc;hn;ht				R					ħ			hT	fc;hc;hn;ht; Q2;b;adapted pureh		hn	hc	ħ		Q2;hT	ht;b		Fc; adapted pureh			
	Award Winner		g;AR;mg;R; hα; gα				Alt h				AR		hα; gα				g;mg		hα						AR;R		gα		g;mg
	Published	m; pureH	IQP;x;Alt h;	mquot; ct; at; POP h; AW;AWCR; AWCRpa	DCI	Price				m; POP h; AW;AWCR; AWCRpa		mquot	IQP;x			ct;at	pureh			ct;at			mquot; AW AWCR	IQP;m		Alt h; POP h; AWCRpa pureh		x	
	Published & cited		f;t;	hmx						DCI						Price	f;t;hmx			Price			f;t				DCI		hmx
	Scientist	e;hg;A;	h2;hpd;w;hm; index seq & mat;	h, n;π; dynamic h				h, hg; A; dynamic h	π		n					index seq & mat	e; hpd;h2; w;hm		n; index seq & mat; π		A			e; h2; w	hpd	hm		h	hg; dynamic h
	Seniority		%HCP; Index age & prod											Index age & Prod	%HCP			Index age & prod						%HCP					
	No definition		hw;Class Dur; hrat;grat;hf					hw					grat		Class Dur		hrat;hf		hf			Class Dur		grat			hw		hrat
Measure	Career		Index Age & prod											Index age & Prod															
	Comparison		Hα; index seq & mat; hf	hn;n, π					π		n		hα			index seq & mat	hn;hf												
	Currency			ct;at;hc		Price										ct;at;Price	hc												
	Distribution	A	ħ					A					ħ																
	Durability		Class Dur												Class Dur														
	Effect	m	Q2;f;t;hT	mquot; AW; AWCR; hT						m;AW; AWCR		mquot				hT	Q2; f;t												
	Excellence	e	%HCP;IQP;h2;b; grat;w;	ht									IQP; grat		%HCP		ht;e;h2;b;w												
	Growth		AR;hpd;R					R			AR						hpd												
	Independence	Pure h; Adapted pureH	fc;hm;Alt h;	POP h; AWCRpa			Alt h	fc		POPh; AWCRpa							fc;hm; pureh; Adapted pureh												
	Quality		hw; gα		DCI			hw		DCI			gα																
	Quality & Quantity		x	h				h					x																
	Rank	hg	g;mg;hrat	dynamic h; hmx				hg; dynamic h									g;mg;hrat; hmx												
Citation	Effect	m	Alt h																										
	Hirsch definition	hg;A	hw;R	h; dynamic h																									
	Influence			π																									
	Impact			POP h;AW: AWCR: AWCRpa	DCI																								
	Performance		AR	n																									
	Popularity			mquot																									
	Quality		IQP; ħ; hα; gα; grat; x																										
	Reward		Index Age & Prod																										
	Idea transfer		Class Dur; %HCP																										
	Use over time		hT; index seq & mat ;	ct;at		Price																							
	No definition	e; pure h; adapted pureH	g;fc;mg;;h2;hpd; Q2;b;hrat;f;t; w;hm;hf;fc	hc;hn;ht; hmx																									

Fig. 14.3 Definitions of concepts and measures in 51 ALIRP

ings of similar researchers. When performing indicator calculations, the use of types of author, resource or publication form that differ from those recommended by the developers of the indicator may in turn affect the performance of the indicator. Consequently, results similar to those demonstrated by the developers cannot be guaranteed. Therefore, if we diverge from the qualities of the author, publication and citations as defined by the developer, what implications will this have for the application and interpretation of ALIRP and for the researcher under evaluation? Is the correlation between the variables and the measure that the ALIRP is purported to produce strengthened or weakened? Are there consequences for our confidence in the reliability of the indicator values?

14.3.4 Mathematical Construction

ALIRP can be percentile-based, depicting the prestige of the researcher's publications based on their position within the citation distribution of their peers. More commonly, ALIRP are based on an assumption of averages.

The percentile of a publication or set of publications is determined by creating a citation frequency distribution for all of the publications in the same year, subject category and document type. The papers are arranged in descending order of citation count, and the percentage of papers at each level of citation is determined, i. e., the percentage of a researcher's papers cited more often than their significant peers as in the *IQP* index and *DCI*. It is also possible to come up with an informed estimate of the researcher's performance if they place lower or higher than expected. As percentiles can be normalized for subject area and time, proportions of publication percentiles can be compared between subjects and periods [14.158]. Further, percentile-based indicators are not dependent on the calculation of the arithmetic mean, which is commonly used in ALIRP, as the arithmetic mean should not be used for skewed data—a typical characteristic of bibliometric data at the individual level of aggregation [14.142]. However, percentile indicators can be strongly biased by different methods of setting the decile limits, defining the peer set for comparison and treating tied papers in the evaluation of the percentage of highly cited publications.

Alternatively, average-based indicators tell us something about the central distribution of citations to publications. At high levels of aggregation, such indicators are appropriate as long as they roughly describe performance patterns. Conflicts arise at low levels of aggregation, when such indicators are used to compare individual researchers on a mathematically fabricated average, creating insecurity in both interpretation and expected performance. Average-based indicators that attempt to indicate an overall citation impact are problematic [14.18, 159], because a single statistic of centrality may not adequately summarize the asymmetries of skewed citation distribution, as some publications will have scored the average number of citations, some will have scored higher and some lower. At this lower level of aggregation, we need precise models, and the requirement for precision often leads to complexity in model design. The concept of averages, estimates of centrality, varies greatly in ALIRP depending on the type of mean estimate used—harmonic (f-index), arithmetic (A-index), geometric (t-index), median (m-index) or mode (h-index)—and importantly, the spread, concentration and proportion of the data used in the model. Needless to say, in the evaluation context, different estimates of the average result in very different snapshots of the average performance of a researcher. The arithmetic mean assumes that the distribution of citations to publications is approximately Gaussian (normal), and as the distribution becomes less Gaussian (either by not being symmetric or due to the presence of outliers), the arithmetic mean becomes a worse description of the distribution. The arithmetic mean returns a higher average value than the geometric mean, as the geometric mean gives larger weight to smaller values in a positively skewed distribution. The geometric mean in turn returns a higher value than the harmonic mean, which is the recommended mean estimate when there are extreme outliers. The harmonic mean in turn returns a higher value than the median. To apply average-based indicators successfully, just as in any other statistical test, the bibliometric data must meet important assumptions of normality. One first needs to determine the distribution of the data and assess whether it approaches a normal distribution before applying indicators based on the chosen mean. Doing this will radically reduce the number of indicators that are appropriate for use. Indicators based on the arithmetic mean, the *CPP* for example, will produce inaccurate results if applied to data that are highly skewed, as the data violate the assumption of normality, which is the essential characteristic needed in the data for this statistical test. Not fulfilling assumptions consequently contaminates the conclusion of the indicator and interpretation of the results. Therefore, indicator developers recommended supplementing such fixed-average indicators with other ALIRP that calculate performance in the tail ends of the distribution, such as the e-index and $\hbar$-index. An alternative, perhaps more intuitive and simpler solution when working with skewed data at the individual-level, is to indicate the average performance by describing the number of citations to papers using the *median*, and reporting the minimum and maximum values to calculate the range or interquartile range where appropriate.

As a researcher's citations and publications increase and decrease over time, they are by no means static. Fixed indicators do not capture this evolution; therefore, testing the stability and robustness of dynamic but mathematically complex indicators such as the *generalized h*, *adapted pure h*, *rational h*, *rational g*, *hT*, *hα* and *gα* can drive indicator design, sometimes superseding the applicability of the indicator in practice. These indicators require the establishment of parameters that represent publication and citation practices in specific fields (note: these parameters are not standardized or consensual, but are experimental) or require special software to compute the indicator, as in the *tapered h*, and *h-sequence and matrices*. The *hw* and *dynamic h*-indicators exemplify Lotkanian informetrics in that they utilize a theoretical approach to bibliometric indicators, building on the Lotka power law of mechanisms of size and frequency. Such dynamic indicators attempt to capture how the distribution between citations and publications can give rise to new phenomena in the interpretation of physical referencing behavior and bibliometric distributions, and hence question how they should be studied rather than how to evaluate researchers [14.160]. They utilize the concepts of size and rank frequency to provide steady patterns of the evolution of papers, for example, in the *h*-core over time, i. e., interpreted as quality papers, indicating that no publication can instantly become a highly cited one, thus implying that in dynamic indicators, the well-known inconsistencies of *h* cannot occur [14.161]. Further, this power law approach is used to develop indicators that use citations as a constant rank-frequency function in the mathematical model to increase the granular precision of scholar rankings, as can be seen in the proposal of the *tapered h-index* [14.162]. These same indicators in tests show their superiority over the *h*-index and other simple indicators by correcting for mathematical consistencies, providing granular comparisons between researchers and embodying the inertia of the objects they are designed to measure. However, their application in practice is severely limited due to their complexity. Simple indicators—of which there are many—may be coarser or even mathematically flawed, but when used wisely, and in combination with one another, they offer great potential for well-rounded author-level bibliometric assessments.

The correlation between publications, citations and effect or impact is not a perfect line. Capturing this relationship in a mathematical model is complicated. The order is not perfect, publications and citations can be concentrated, and the distribution can be skewed. Therefore, applying the wrong mathematics is still wrong, even if the methods are well known, as in *CPP* and *h*. Different weights and parameters included in the indicator model may lead to different conclusions about researcher performance. Indicator values can be highly affected by the researcher's publishing and citing behavior, or based on a system of proportional representation where such behavior is not considered important, leading to a loss of detail.

All models can lead to mistaken observations and conclusions about the performance of the researcher within the boundaries of the evaluation system, but understanding the math and applying the mathematical models to data that fit the underlying assumptions will greatly improve the validity of our estimates of researcher performance.

14.3.5 Families of Indicators

ALIRP are a specific species of bibliometric indicator, and because this chapter advocates the application context, families of ALIRP are presented in the following taxonomy (Fig. 14.4) based on what they are designed to measure rather than a genus taxon based on math. The order of the families in the taxonomy depicts the increase in indicator complexity, reading from left to right, the simplest counting models closest to the raw output [14.2]. There are no hard and fast rules for describing or recognizing a family, branches or any taxon. One could take different positions regarding taxonomic descriptions, and I do not claim that Fig. 14.4 is *the* correct depiction. Figure 14.4 increases our awareness of the families and relationships in the community of ALIRP.

Starting with the light blue family, these indicators count a researcher's output. These are models of how to count different types of scholarly and scientific works, published or unpublished depending on the unit of assessment, adjusting for publication channel, bibliographic database and co-authorship. These are closely linked to the green family of indicators that measure the citation effect of the researcher's output by counting whole citations or by counting fractional citations to adjust for collaboration behavior. This family contains two main branches of indicators of the citation effect of a researcher's output. In the first branch, citations are normalized to publication type or a specific field to enable comparison with standardized benchmarks. On the second branch, citations are normalized to certain publications within the researcher's portfolio of work or to the age of the citation to identify work that is actively being used. Indicators that qualify output using journal impact factors are depicted in yellow. In this family, the first branch of indicators measure the impact of a re-

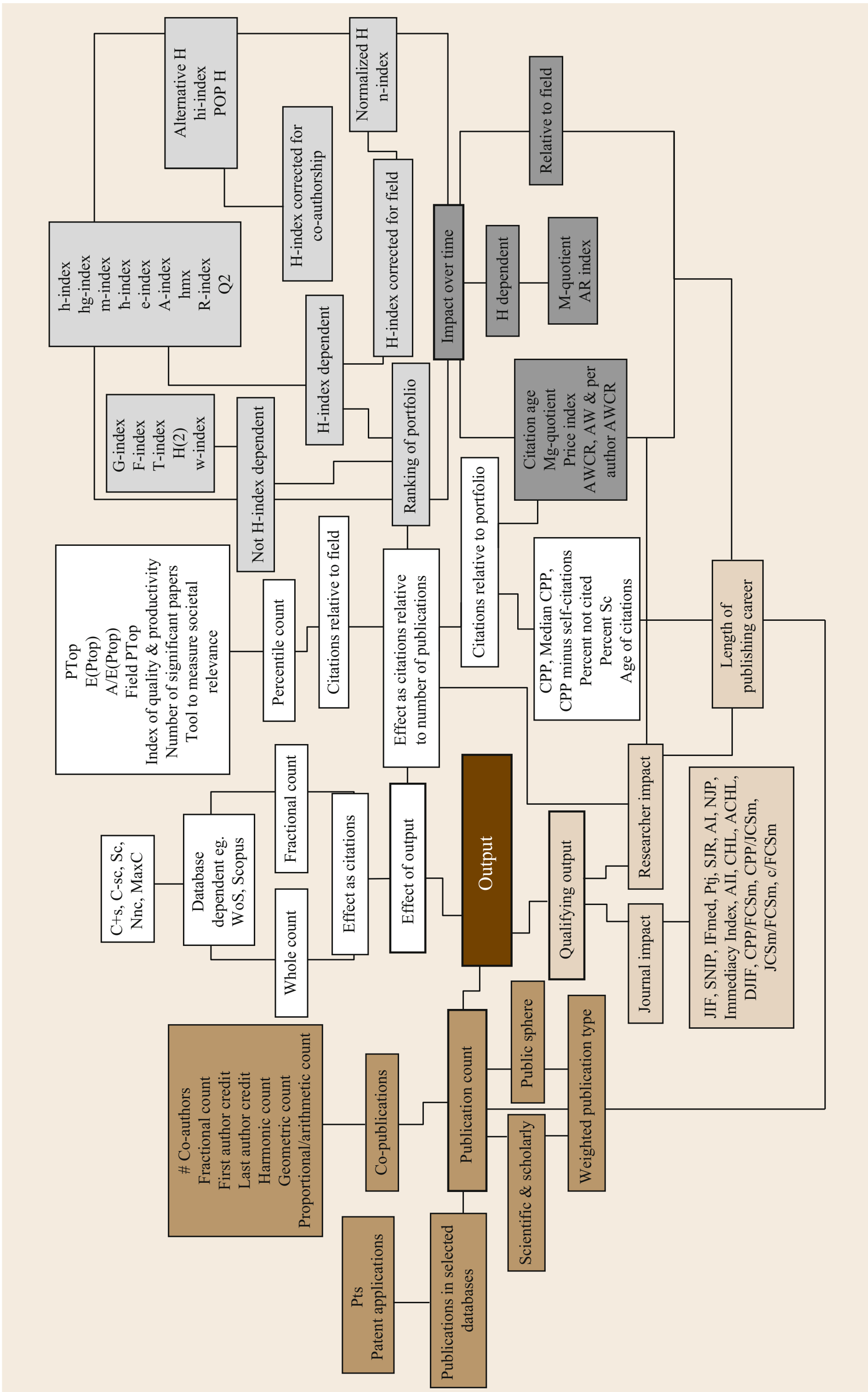


Fig. 14.4 Taxonomy of 51 ALIRP

searcher's articles in chosen journals, where the impact of the journals themselves is used to formally suggest the potential visibility of the researcher's work in the field in which he/she is active. The second branch, researcher impact, suggest the impact of the researcher's portfolio of work by using a combination of indicators from the output and effect families to formally suggest the researcher's productivity and visibility of work in the field in which he/she is active. Typically, these indicators are used in combination with a narrative, e. g., *IQP*, *index of age and productivity*, and the *Classification of Durability*, to situate the indicator values in the context of the researcher. In this family, indicators qualify output by normalizing for the length of a researcher's publishing career. The purple family of indicators, on the far right, contains indicators that measure and rank the portfolio of the individual, typically to promote the top-performing publications, or in an evaluation, to promote the researchers themselves. These indicators rank publications by the number of citations each publication has received and establish a mathematical cutoff point for what is and is not included in the ranking. The indicators are subdivided into two branches, h-dependent and h-independent. The h-dependent indicators include the calculation, or a variation thereof, of the h-index in their construction in order to denote something about the importance or significance of a researcher's work or to rank researchers. Additionally, the indicators can adjust h to account for the characteristics of research areas or co-authorship. Secondly, there are the h-independent indicators that do not include h in their construction. They provide an alternative to h, attempting to avoid the flaws of h while still producing values that denote something about the importance or significance of a researcher's work or that rank researchers, adjusted to field or co-authorship. Finally, the orange family of indicators of impact over time denote something about the extent a researcher's output continues to be used or the decline in use. This family includes indicators that set the researcher in the context of his or her research area (relative to the field) and include h-type indicators that identify the use of the active core of the researcher's work over time.

Figure 14.1 illustrates quite simply the many groups of indicators we have at our disposal, those that share common characteristics and where they differ. We can now begin to get an idea of which indicators could complement and supplement one another in an extensive analysis that could capture the diversity of the production and performance of a researcher's output.

We still have much to learn about the extent to which ALIRP supplement one another in practice, whether they become redundant when used together, produce too much or conflicting information, or whether they do indeed contribute with unique information that tells us something highly individual about the researcher's output and the extent of its use. In application contexts, the first step is to understand the conceptual design and mathematical construction of each applied indicator. This will help us discern what and how we measure, and whether we are indeed measuring what we value, or are just measuring what we can easily measure and thus valuing what we (can) measure [14.163]. Like any other mathematical model, ALIRP are dependent on the quality and configuration of the data used to compute them as well as our expertise in interpreting the model. Different weights and parameters included in the indicator model may lead us to different conclusions about the researcher's performance. Selecting appropriate indicators involves making explicit choices about how to account for diversity. One indicator alone cannot possibly encompass all the requirements of an assessment or capture the similarity, disparity and balance between researchers and the objects being measured. Determining the most appropriate indicators to use in practice depends on the goal of the evaluation and the context of the researcher. Therefore, methodological transparency is essential. This includes describing the data and sources used in the calculation of the indicators as well as an interpretation of the math to show the properties of what you are measuring, i. e., how you are using the model to provide an interpretation of a researcher's production and effect. It is quite acceptable to use different models and different math in evaluations, as these result in different perspectives of interpretation of a researcher's performance.

In summary, this section argues that the choice of indicator must be substantiated through critical reflection on the conceptual design and mathematical construction of the indicator models before application in practice. The rationale is simple: to provide valid answers to questions about the production and use of a researcher's work, we need good measures. To better understand the quality of the measures we have, this chapter began with a description of the evolution of ALIRP and a discussion of the general characteristics of ALIRP. The discussion included a description of the mathematical and theoretical foundations of ALIRP, their conceptual clarity, validity and intended purpose, and their strengths and weaknesses. These discussions culminate in the following section in the form of three extensive tables that explicitly define the design, mathematical model, advantages and limitations of ALIRP. The purpose of the tables is to provide an overview of the ALIRP discussed above and to enable consumers of ALIRP in evaluation processes to assess the appropriateness of the indicators and consider whether the use of the indicator is valid in an evaluation context.

14.4 Schematizing the Indicators

In the following sections, the common features of 69 indicators designed for use at the individual level are summarized in table form. The aim is to illustrate the diversity of the indicators and the potential we have for multidimensional bibliometric analyses.

14.4.1 Introducing the Tables

Table 14.1 presents indicators that count publications, Table 14.2 provides indicators that count citations, and Table 14.3 lists hybrid indicators (offspring of publication and citations counts that combine elements of publication and citation performance with other components and mathematical manipulations). In total, 69 suggested ALIRP are schematized with regard to six criteria, which appear as column headings in the tables. The indicator name/acronym is presented in the first column, followed by the concept the indicator is design to capture in the next. In the third column, a brief description of how to compute the indicator is introduced. The fourth column contains the definition of the indicator, which is taken from the developers' own description of the purpose of the indicator in their original published papers. This is followed in the next two columns by a brief presentation of the advantages and disadvantages of the indicator as discussed in the bibliometric literature, supplemented by additional comments in the final column. The bibliographic references to each indicator, where available, are found in an additional reference list in the appendix of this chapter. The 69 indicators in Tables 14.1–14.3 are systematically assessed in order to better understand the unique approach of each indicator and its suggested use in evaluating an individual researcher. In this respect, the tables are designed to support an informed choice across a range of ALIRP by examining the strengths and weaknesses in indicator construction that can enhance or hinder their effective use in evaluation.

With no advisory boards, common standards or contextual assessments, comparing indicators is problematic, and this can both impede the development of the field and cause consumers of ALIRP to be distrustful of the results of evaluations [14.49]. On the other hand, without the restrictions of advisory boards, common standards or contextual assessments, creativity in indicator development is unfettered. Professor *Yves Gingras* addressed this duality in 2014, and suggested that developers, administrators and researchers need to learn how to better evaluate indicators [14.157]. He provides three validity criteria:

(C1) The adequacy of the indicator for the object/concept it measures

(C2) The sensitivity to the intrinsic inertia of the object/concept being measured, i. e., that there are no important changes without cause and that the indicator increases proportionally in relation to the concept

(C3) The homogeneity of the dimensions of the indicator, e. g., ensuring that the units in the mathematical model are consistent on both sides of the equal sign, as in $CPP = C/P$, and that the indicator is not combined with arbitrary weights. Homogeneity can tell us whether the math is incorrect, but cannot tell us if the math is definitely correct.

It is important to remember that even ALIRP that fulfill all three criteria could still produce invalid results in an application context if the scope or precision of the data collection process is flawed.

Each of the 69 indicators, presented in the tables in the following sections, was assessed using the validity criteria. This assessment adhered to the guidance for critical reflection provided by *Gingras* [14.157], which involved reviewing the literature about the construction of each indicator (summarized in previous chapters and illustrated in Fig. 14.3) and learning about the strengths and weakness of the indicator in practical applications (summarized in Tables 14.1–14.3). This approach has resulted in a qualitative analysis attempting to determine the exact meaning of each indicator. In the text supplementing the tables of indicators, the results of this analysis are presented not as the infallible truth, but to spark the reader's own critical reflections on the potential validity of the indicators before applying them in practice.

All three tables may be downloaded as Excel files that are made available together with the online version of this chapter. The tables are extensive, detailing the evaluation criteria and the complexity of each indicator.

14.4.2 Indicators that Count Publications

Ten common publication indicators are presented in Table 14.1. The strength of publication count models lies in their easy adaptability by consumers to count only publications from specific sources or to include the type(s) of publications deemed important for the evaluation or discipline. This can be done without compromising the properties of the specific indicator. *Cognitive orientation*, *publications in selected sources*, *P* and *weighted publication count* are examples of the aforementioned and are indicators that clearly correspond to the concept they are defined to measure (C1). They

Table 14.1 Indicators that count publications

Indicator	Measure	Computation	Definition	Advantages	Disadvantages	Comment
Cognitive orientation	Cognitive orientation	Aggregates papers in scientific subfields in which the individual publishes or is cited	How frequently a scientist publishes, is cited in various fields	Relatable to the position a researcher holds in the community	Limited to the database definitions of scientific fields. Journal-based, more applicable in some fields than others	Useful for identifying future areas for collaboration and production
App proportional	Contribution	Author with rank R in a byline with n co-authors ($R = 1 \ldots n$) receives score $n + 1 - R$	Shared authorship of papers, first author weighted highest and last author lowest	Rewards level of contribution to a paper	Indicator cannot be used if authors listed alphabetically or where authors alternate first authorship	Can be normalized in such a way that the total score of all authors is equal to 1
App geometric	Contribution	Author with rank R with N co-authors receives credit of $2N - R$	Assumes that the rank of authors in the byline accurately reflects their contribution	The first few authors receive most of the credit	Allotted authorship credit rapidly approximates asymptotic values as N increases	Asymptotic values lose their validity with small sample size
App harmonic	Contribution	Ratio of credit allotted to i-th and j-th authors is $j : i$ regardless of total number of co-authors	First author gets twice as much credit as the 2nd, who gets 1.5 times that of the 3rd, who gets 1.33 times that of the 4th etc.	Representation of perceived quantitative norms of byline hierarchy	Applies only in areas where unequal co-author contributions are the norm	Tested in the natural sciences
FA	Contribution	Only first of n authors of a paper receive credit equal to 1	Credit given to first author only	Simple method of crediting publication to the assumed main contributor	Does not give an accurate picture of the relative contribution of the authors	Unfair when authors are ordered alphabetically or practice *noblesse oblige*
Noblesse oblige	Contribution	Last author receives 0.5 credit, other $N - 1$ authors receive $1/(2(n-1))$ each	Indicates the importance of the last author for the project behind the paper	Last author gets credit for contributing with resources (funding, network, expertise) and ensuring the article was written	There is no way to identify actual level of contribution apart from authors' statements	This is one of many suggested counting schemes for noblesse oblige
Fp	Independence	Each publication divided by number of authors, limited to max. 10 authors	Shared authorship of papers gives less weight to collaborative works than non-collaborative ones	Accounts for differences in publishing behavior among fields and level of multi-authorship	Favors secondary authors by allocating equal credit to all authors	Criticized for lack of fit between credit scores and contribution
Publications in selected sources	Production in selected sources	Count of publications in predefined sources	Important sources defined by scholars, institute, field or evaluation committee	Reflects output in sources deemed locally important	Provides a distorted or incomplete picture of production	Provides only a snapshot of productivity
P	Production of papers in journals and academic books	Sum of publications	Count of production used in formal communication	Potentially, all types of output can be included or selected with regard to theme of evaluation	Does not measure importance, impact of papers, duration or volume of research work	Counts vary across disciplines due to nature of work and conventions for research communication
Weighted publication count	Production of specific types of publications	Applies a weighted score to the type of output	Distinction between different document types	Importance of different publication types for communication within a field	Has to be designed for individual specialty/field	Enables comparison of like with like if agreement of importance is reached

are homogeneous indicators (C3) and fulfill the inertia criteria (C2), that is, the indicator value increases in a manner consistent with an increase in the concept that the indicator measures. All 10 indicators are simple to calculate and make sense; however, caution is advised when interpreting the *Fp*, *noblesse oblige*, *FA*, *App harmonic*, *App geometric* and *App proportional* indicators. These indicators work with the concept of contribution, that is, the substantive intellectual contribution to a paper an author has made can be mathematically deduced from his or her position on the author byline of a paper. There is no way to identify the actual level of contribution apart from gathering statements of authorship. These indicators instead offer a complementary insight into the impact of an author, adjusting for potential biases of (excessive) co-authorship. As a set, the publication indicators presented in Table 14.1 supplement one another, in that they provide unique insights into the quantity of publications an individual has produced.

14.4.3 Indicators that Count Citations

Eight citation-counting models are presented in Table 14.2. Five are considered to correspond to the concept they are defined to measure (C1), fulfill the inertia criteria (C2) and are homogeneous (C3). They are as follows: *C*, *database-dependent citation count*, *C-sc*, *%nc* and *%sc*. These indicators are also simple to calculate, and what they measure is unambiguously clear. They are independent indicators, in that they provide information about different aspects of a researcher's citation count. The indicators *SIG*, *Sum pp top prop* and *Fc* have obvious flaws in their conception, but because they are so intuitively simple, they could still be appropriate measures for evaluation. *Fc* attempts to give less credit to collaborative works, leading to fairer evaluative comparisons. However, credit is calculated in *Fc* as a single unit that can be distributed evenly, making the share dependent on the number of authors, thus assuming that the contribution of the author can be equally rewarded. *SIG* and *Sum pp top prop* are coarse indicators of significant works, and these significant works are defined as top papers in the context of a database's indexing policy and classification system. When interpreting these indicators, it is wise to draw the consumer's attention to that fact that the paper with the greatest number of citations is not necessarily a researcher's most academically/scientifically significant paper. Further, the indexing policies and classification systems in databases can have both geographic and disciplinary biases and resource bias, which further distorts the interpretation of the indicator [14.10, 164].

14.4.4 Hybrid Indicators

Fifty-one hybrid indicators are analyzed in Table 14.3. Ten of these are identified as relatively simple to compute, correspond to the concept they are defined to measure (C1), fulfill the inertia criteria (C2) and are homogeneous (C3). They are as follows: f, t, *index of age and productivity*, π, $c(t)$, $a(t)$, *Price index*, *Classification of Durability*, *AWCR*, and *IQP*. The commonality of the these 10 ALIRP is that they cannot stand alone. They all demand a narrative to set the indicator values in the context of the researcher and of the calculation—the *IQP* is an excellent example. The narrative is a brief description showcasing the informed use and informed interpretation of the bibliometric indicators, and the expertise and influence of the researcher in the context of their demographic information, specialty and academic seniority. These indicators have been tested empirically in WoS and other databases and are recommended as suitable for implementation in different disciplines (for more information, please refer to the bibliographic references to these indicators at the end of this chapter). Further, there is no redundancy among these 10 indicators when they are used together, as each indicates different facets of the citation effect of a researcher's publications. They can be used together, within the context of the citation index, to indicate:

- The currency and relative currency of the researcher's work ($c(t)$, $a(t)$, *Price*, *Classification of Durability*)
- The average effect of a researcher's work (f, t)
- The cumulative effect of the researcher's body of work, (*AWCR*)
- The excellence of the researcher's work compared to subjectively defined specialty standards that are based on the researcher's publication habits (*IQP*, *Index of Age and Productivity*).
- Comparison with peers, across fields (π).

The remaining 41 ALIRP fail on one or more of Gingras' evaluation criteria (please refer to the tables provided as e-material to this chapter for more detail). Thirty of these indices are corrections to the h-index, using the principles of h in their calculation. Eight are alternatives to the h index that do not use h in their calculation, and three are diverse ALIRP that indicate the average number of citations to all papers, top papers and the currency of a researcher's work: the *CPP, %HCP* and *DCI*.

The 38 h-type ALIRP attempt to correct for the inherent inconsistency problem in h, in that h performs in rankings in a counterintuitive way [14.135]. With consistent indicators, it is ensured that if two authors

Table 14.2 Indicators that count citations

Indicator	Measure	Computation	Definition	Advantages	Disadvantages	Comment
C	Times cited	Sum of citations, including self-citations	Number of citations for whole period of analysis	Reflects social side of research and cumulative growth of knowledge	Quality and timeliness of citation not considered	Self-citations affect the reliability and validity of the measure on small amounts of data in assessments
Database-dependent citation counting	Times cited in specific sources	Number of citations recorded in a specific database	Citation number is dependent on the database used to collect citation information	Shows that coverage of researcher in database can affect calculation of indicators and performance of researcher	Many indicators and field benchmarks are reliant on WoS and cannot be compared with data from other sources	Scope, validity, reliability and cost of the citation collection is dependent on choice of citation index
C-sc	External citations	Sum of citations, minus self-citations	Measure of external citations	Reflects social side of research and the cumulative growth of knowledge	Quality and timeliness of citation not considered; what is a sc: cites of oneself, a co-author or institutional colleague	Does not account for older articles being more cited and variation in citation rates between document types and fields
Sig	Most significant work	The paper with the highest number of citations	The most significant paper	Simple indicator of the most important and influential research	The most significant paper is not necessarily the paper with the most citations	Most highly cited paper can have the largest number of authors, be longer than the average article and have more references
Sum pp top prop	Most significant work	Proportion of papers in the top 10% in the world	Identify scholars' papers that are rated top of their field	Simple indicator of most cited papers in the field	Uses WoS subject categories to define field	Does not compare like with like
%nc	Work that has not been cited	The number of uncited papers divided by sum of citations, multiplied by 100	Percentage of papers that have not been cited	Used to contextualize other indices, e. g., CPP, related academic age	Papers may not be recorded as cited due to citation database indexing policy	Illustrates the types of publications that, although important, do not receive citations, i. e., technical reports, guidelines
%sc	Self-use	Self-citations divided by sum of citations multiplied by 100	Percentage of self-citations	Illustrates how work builds on previous findings. Advertises the work and the author	Unclear what a self-citation is: cites of oneself, a co-author or institutional colleague	Self-citation rate is highly variable among individuals and their contribution highly variable
Fc	Citations scholar would have received if worked alone	Citations to a paper divided by the number of authors. Limited to max. 10 authors	Removes the dependence of co-authorship on citation count	Less weight given to collaborative works	Regards credit as a single unit that can be distributed evenly across a number of authors	Comparison to field norm unwise, as citations to the publications may not be representative of the field, but biased toward the highly or poorly cited

Table 14.3 Hybrid indicators

Indicator	Measure	Computation	Definition	Advantages	Disadvantages	Comment
f	Average effect	Harmonic mean of citations to papers	Highest number of articles that receive f or more citations on average	An additional citation to an infrequently cited paper counts more than an additional citation to an often-cited paper	Both f & t indices are maximum if every paper is cited the same number of times. The f-index deviates much faster from this maximum than the t-index	More discriminatory power than h- and g-indices, as the nonlinearity of the harmonic mean is more sensitive to small differences between researchers
t	Average effect	Geometric mean of citations to papers	Highest number of articles that received t or more citations on average	Geometric mean does not place much weight on the distribution of citations	Sensitivity to small differences between researchers is stronger with harmonic mean (f-index) than geometric mean	It is not sufficient to determine the function and value of citations using indices; their cognitive background should also be taken into consideration
CPP (CPAY)	Average number of citations per paper	Sum of citations divided by sum of papers (alt. number of years since first publication registered in citation index)	Trend of how cites evolve over time	Enables comparisons of scientists of different ages and different types of publications	Tells nothing of the timeliness, origin or quality of the cite (positive or negative)	Citations can be hard to find, reward low productivity & penalize high productivity
w-index	Broad impact of masterpieces	w is the highest number of papers that have at least 10w citations each	Impact of researcher's most excellent papers	Accurately reflects the influence of a scientist's top papers	Tendency to describe quantity of the productive core	w-index of 1 or 2 = someone has learned the rudiments of a subject; 3 or 4 = mastered the art of scientific activity, *outstanding individuals* have w-index of 10
Index of age & productivity	Career	Mean number of documents by age and CPP (3-yr citation window) in 4-year age brackets, adjusted to field	Effect of academic age on productivity and impact	Identifies the age at which scholars produce their best work and the extent of the decline of their production	Mean impact declines with age regardless of quality of researcher's body of work	If used independently, fosters practice of quantity over quality. Difficult to maintain high values of impact with increasing rates of production
Hn (Normalized h)	Comparison across fields	hn = h/papers(Np), if h of its Np papers have at least h citations each, and the rest (Np − h) articles received no more than h citations	Normalizes h to compare scientists achievement across fields	Accounts for scientists different publication and citation habits in different fields.	Hn can only be used with h-index; rewards less productive but highly cited authors	Gives paradoxical results for junior researchers with few papers

Table 14.3 (continued)

Indicator	Measure	Computation	Definition	Advantages	Disadvantages	Comment
π index	Comparison across similar fields	π is one hundredth of the number of citations received by the top square root of the total number of papers ranked by decreasing number of citations	Production and impact of scholar	Comparative assessment of scientists active in similar subject fields. Sensitive to citedness of top papers	Value depends on citation rate of papers in the elite set (top-cited papers); the elite set is scaled by an arbitrary prefactor	Can be calculated on a small number of papers. Unique index as it is defined in terms of the summed number of citations rather than the square root of the sum or the average
hf	Comparison to peers	Citations of each article normalized by average number of citations per article in the subject category of the article under observation	Comparison to peers by correcting each article's citation rate for field variation	Makes scientists of different scientific age comparable	Difficult to determine the correct publication/citation window in construction of the matrix	Calculation is not easy, making it a nominal index and not a pragmatic one
H index of sequences and matrices	Comparison to peers and domain	Calculates h-sequence by continually changing the time spans of the data. Constructs h-matrix based on a group of correlative h-sequences	Identifies significant variations in individual scientists' citation patterns across different research domains	Makes scientists of different scientific ages comparable	Difficult to determine the correct publication/citation window in construction of the matrix	Only tested on 11 well-established physicists
n index	Comparison within specialty & across fields	Researcher's h-index divided by the highest h-index of the journals of his/her major field of study	Enables comparison of researchers working in different fields	Can surmount the problem of unequal citations in different fields	Still awaiting validation	Calculation based on Scopus definition of h and SCImago Journal and Country Rank website for journal information
c(t)	Currency	c(t) is the difference between the date of publication of a researcher's work and the age of citations referring to it	The age of citations referring to a scholar's work	The entire distribution of the citation ages of a set of citing publications provides insight into the level of obsolescence or sustainability	Measuring aging by means of citation counting is disputed. Does not account for effect of literature growth, availability of literature and disciplinary variety	Usage and validity are not necessarily related
a(t)	Currency	a(t) is the difference between ct and $c(t+1)$	Aging rate of a publication	Allows for translation of diverse factors influencing aging into parameters that can be estimated from empirical data with a specified margin of error	A corrective factor is required if citation rates are to be adjusted for changes in the size of citing population and discipline	For individual documents, stochastic models are preferable. The simplest model to study aging is the exponential decay of the distribution of citations to a set of documents

Table 14.3 (continued)

Indicator	Measure	Computation	Definition	Advantages	Disadvantages	Comment
Hc	Currency	Articles assigned decaying weight depending on age. The weighting is parameterized gamma = 4 and delta = 1	Currency of articles in the h-core	Accounts for active versus inactive researchers	An old article gradually loses its "value", even if it still receives citations; thus newer articles are prioritized in the count	Identical to hpd, except measured on a 4-year cycle rather than a decade
Price index	Currency	PI = (n1/n2) × 100. n1 = number of cited references with a relative age of less than 5 years. n2 = total number of references	Percentage references to documents, not older than 5 yrs, at time of publication of citing sources	Accounts for differing levels of immediacy characteristic of the structurally diverse modes of knowledge production in the different sciences	Does not reflect the age structure in slowly aging literature	In calculation of PI it is unclear whether the year of publication is year 0 or year 1. It is unclear whether this year is included
A	Distribution of citations	Arithmetic average number of citations to articles in the h-index	Magnitude of scholars' citations to publications, supplement to h	A-index can increase with increase in citations, even if h-index remains the same	A is h-dependent and has information redundancy with h. When used together, h masks the real differences in excess citations of different researchers	A-index involves division by h and punishes researchers with high h-index; sensitive to highly cited papers
$\hbar$	Distribution of citations	Square root of half the number of total citations to all publications	Overall structure of citations to papers. Enables comparison across field and seniority	Includes papers h ignores, i. e., most highly cited and the body of articles with moderate citations	Difficult to establish the total citation count with high precision	$\hbar$ is only roughly proportional to h
Classification of Durability	Durability	Percentile distribution yearly citations to a document, accounts for all document types and research fields	Durability of scientific literature	Aids study of individuals from general perspective using composite indicators. Discriminates between normal, flash in the pan and delayed publications	Minimum 5-year citation history threshold for reliable results. Empirically investigated in WoS using journal subject categories	Can be applied to large sets of documents or documents published in different years; documents can be classified in more than one field and can be updated yearly or monthly
hT index	Effect of all papers	Using Ferrers graph, h-index is calculated as equal to the length of the side of the Durfee square assigning no credit to all points that fall outside	Production and impact index that takes all citations into account, yet the contribution of the h-core is not changed	Evaluates the complete production of the researcher, all citations, giving to each of them a value equal to the inverse of the increment that is supposed to increase the h-index one unit	The relative influence of the interpolation will be stronger for smaller values of the indices; therefore, utilize the generalized indices when comparing many data sets with very small values of h	Shows smooth increase in citations, not irregular jumps as in h-index. Conceptually complex

Table 14.3 (continued)

Indicator	Measure	Computation	Definition	Advantages	Disadvantages	Comment
Q2	Effect of all papers	Square root of (h multiplied by median number of citations to papers in h index)	Relates two dimensions in productive core: the number and impact of papers	h-index' measurement of papers in core, correcting for distribution of citations to papers	h- and m-index have to be obtained before calculation of Q2	Geometric mean is not influenced by extreme higher values. Fuses the information provided by the aggregated values in a balanced way
AWCR	Effect of all papers	(Citations to all papers, divided by age of paper)/number of publications	AWCR = number of citations to all papers, adjusted for age of each individual paper	Sum over all papers represents the impact of the total body of work. Younger, less cited papers contribute to AWCR	Field norm has to be decided to account for field characteristics, e. g., expected age of citations, "sleeping beauties" and delayed recognition	AWCR offers by default empirical insight into research seniority and career age
AW	Effect of all papers	Square root of AWCR	Square root of AWCR avoids punishing researchers with few very highly cited papers	AW approximates the h-index if the mean citation rate remains constant over the years. Includes younger and yet less cited papers	It is more rigorous to assign weights to each of the publications, calculate the average weighted citations (arithmetic mean), and normalize that result to one of the publications	Square roots are appropriate for data that grows exponentially. Gives consistent results, though this is not the same as giving correct results
m-quotient	Effect of best papers	h-index divided by years since first publication	h normalized for academic age	Comparisons between academics with different career length, as h is approximately proportional to career length	m stabilizes later in career; small changes in h can lead to large changes in m; first paper not always an appropriate starting point	m-quotient adds time as a weighting factor. It does not counter the major disadvantages of the h-index
mg-quotient	Effect of best papers	g-index divided by years since first publication	g normalized for academic age	Comparisons between academics with different career length. g is approximately proportional to career length	First publication is not necessarily the appropriate estimate of the start of the researcher's career	mg and m discriminate against part-time researchers/career interruptions
m	Effect of best papers	Median citations to publications included in h-index	Reduces impact of highly cited papers	Accounts for skewed citation distribution. Median is used as measure of central tendency	Median may be a better measure of central tendency, it can be chronologically unstable	Reduces impact of heavily cited papers
%HCP	Excellence	Publications cited above the 80th percentile in respective research areas	Indicates papers among the 20% most cited in research area, i. e., relative impact	Indicator of excellence understood as citation count reflect the extent to which an academic's work affects the work of his/her peers	Indicates only one facet of excellence and no reflection of the impact of the work on society	Difficult to maintain high values of relative impact with increasing rates of production

Table 14.3 (continued)

Indicator	Measure	Computation	Definition	Advantages	Disadvantages	Comment
IQP	Excellence	A = (mnjs × Pyrs x p + 1)/2. B = A/number of papers C = Received citations IQP = C/B + nPapers, C = calculate field impact per papers x number of papers	Expected average performance within field; number of papers cited more frequently than average and how much more these papers are cited	Corrects citation count for scholarly productivity, author's academic age, and field-specific citation habits. IQP Calculator: [14.165]	Tested in natural sciences, medicine and psychology. Dependent on WoS field specific journal impact factors	Correlates better with expert ratings of excellence than h index. Puts papers in low cited fields on same scale as papers in highly cited field to improve comparison
e	Excellence	Defines total citations to articles in h-index. Subtract h^2 from total citations (e^2). Square root of e^2 is e	Production and effect of highly cited papers	The combination h and e provides complete citation information	e can only be calculated if h is given	More precise in evaluation of highly cited scientists and comparing group of scientists who have identical h
H2	Excellence	Cube root of sum of citations	Cumulative achievement. Weights most productive papers	Reduces precision problem as only a small subset of the researcher's papers used to calculate H2	Difficult to discriminate high H2 indices between scientists with different numbers of publications and different citation rates	Suffers from same intrinsic disadvantages as other indices based on citations, i. e., it is not comparable across fields or academic age
b index	Excellence	b = integer value of the author's external citation rate (non-self-citations) to the power three quarters, multiplied by h	Number of papers i that belong to the top n% of papers in a field; Effect of self-citations on h	Field-specific reference standard for scientific excellence used as cutoff value for including or excluding publications in productive core	Difficult to implement because of the computations needed to obtain the measure. Difficulty in obtaining accurate data from bibliographic databases	The b index depends on the year in which it is determined, the period under consideration and the database
Rational g	Excellence	Interpolates between g and g + 1 based on the piecewise linearly interpolated citation curve	Indicates the distance to a higher g-index	It is not a complementary index requiring first the determination of g, but rather follows from a self-consistent definition	Relative influence of the interpolation is stronger for smaller values of the indices. Utilize generalized indices when comparing many data sets with small h values	As every citation increases interpolated g, the index is sensitive to self-citations
AR	Growth	Square root of sum of the average number of citations per year of articles in the h-core	Citation intensity and age of articles in the h-core,	AR is necessary to evaluate performance changes. Supplements h	The decay of a publication is very steep and insensitive to disciplinary differences	AR increases and decreases over time. AR is not convincing as a ranking metric in research evaluation

Table 14.3 (continued)

Indicator	Measure	Computation	Definition	Advantages	Disadvantages	Comment
Hpd	Growth	Hpd is highest number of papers that have at least hpd citations per decade each and other papers have less than hpd + 1 citations per decade each	Compares the output of scholars of different ages. Seniority-independent h-type index	hpd of a mature scientist is nearly constant over many years, and hpd of an inactive scientist slowly declines whereas h-index increases in time	Using scaling factor of 10 to improve granularity between researchers is an arbitrary number. Randomly favors or disfavors individuals	Can be modified for multi-authored papers. A paper's cites per year is divided by the number of co-authors to indicate contribution of each co-author
R	Growth	Square root of the A-index	Magnitude of scholars' citations to publications	Adjusts for punishing the researcher with a high h-index	R-index involves division by h and punishes researchers with high h-index	Supplement to h. Easier to calculate than g index, but not as elegant
Hm	Independence	Hm is the reduced number of papers that have been cited hm or more times	Softens influence of authors in multi-authored papers	Does not push articles out of the h-core; each paper is fully counted allowing for a straightforward aggregation of data sets	Precision problem is enhanced, as additional papers enter into the hm-core	Uses fractional paper counts instead of reduced citation counts. Uses inverse number of authors to yield a reduced or effective rank
Alternative h	Independence	Alternative h = h-index divided by mean number of authors in the h publications	Number of papers a researcher would have written along his/her career if working alone	Rewards scientists who work alone from authors that work in groups. Accounts for differences in co-authorship patterns, discipline and self-citations	Mean is sensitive to extreme values and could penalize authors with papers with a large number of authors	Might decrease when a paper with many authors advances into the h-core by attracting additional citations and reduces size of the h-core
POP h	Independence	Divides number of citations by number of authors for that paper. Normalized citation counts used to calculate h	Accounts for co-authorship effects	Gives an approximation of the per-author impact, which is what the original h-index set out to provide	Normalization by mean number of authors of publications in the h-core leads to reduction of the index	Also considered multiple authors by computing g and h indices using a fractional crediting system
AWCRpa	Independence	Citations to a paper/age of paper. Result for each paper divided by the number of authors per paper. Summed over all papers	Normalizes AWCR to the number of authors for each paper	Accounts for co-authorship to give a more approximate individual author impact	Giving less weight to collaborative publications than non-collaborative publications is undesirable	Rankings using AWCRpa agree with those that rely heavily on citations per year data
Pure H	Independence	Hp = square root of h/normalized number of authors and credit to their relative rank on the by-line of h-core articles	Corrects individual h-scores for number of co-authors	Reduces effect of collaboration in multi-authored, highly cited paper	Results vary depending on method of distributing credit to authors—fractional count, arithmetic to determine h	More refined index is to use the number of collaborators, rank in the byline and the actual number of citations (pure R)

Table 14.3 (continued)

Indicator	Measure	Computation	Definition	Advantages	Disadvantages	Comment
Adapted pure h	Independence	h is interpolated rank value between papers (fractionally counted) and citations (counted as square root of equivalent number of authors)	Finer granularity of individual h-scores for number of co-authors by using a new h-core	Alters h-core to be less biased than Hp with respect to authors with many multi-authored papers	Precision is an issue and difficult to calculate	Leads to a more moderate correction of authorship than h_i as divides citation count by the square root of author count rather than full author count
Ht	Pioneer research	Each citation of an article is assigned an exponentially decaying weight, which is expressed as a function of the *age* of the citation	Age of article and age of citations	Identifies pioneering articles that set out new line of research and still cited frequently. For researchers with a long career behind them	The weighting is parameterized, and for gamma = 1 and delta = 0, this metric is the same as the h-index	Estimates impact of researchers' work in a particular time instance, i. e., whether articles still receive citations by looking at the age of the cites
hw	Quality	Hw = square root of total weighted citations (Sw) received by the highest number of articles that received Sw/h or more citations	Quality of a scholar's publications (h value weighted by citation impact)	Improves sensitivity to the number of citations in h-core	Does not use h-table in calculation and is therefore not an acceptable h-type measure	Hw can be misleading and a contradiction of h
DCI index	Quality	Sum of weighted count of citations over time to a set of documents divided by the logarithm of the impact in past time intervals	Devalues old citations in a smooth and parameterizable way	Gives more weight to highly cited publications as these are assumed to be quality works. Weights citations by the value of the citing publication	Difference caused by weighting: some authors gain impact while others lose	Rewards an author for receiving new citations even if the publication is old
h	Quality and quantity	Publications ranked in descending order after number of citations. Where number of citations and rank match, this is h	Cumulative achievement (productivity and impact)	h is a simple but rough measurement of quality of work	Once a paper is in h-core, the number of citations it receives is disregarded. Loss of citation information means comparisons based on the h-index can be misleading	Arbitrary cutoff value for including or excluding publications from productive h-core
x index	Quantity and quality in cross disciplinary comparisons	x is a researcher's absolute score divided by a reference score	Describes quantity and quality of the productive core and comparison with peers	Uses the journals the researcher publishes in as reference, not field classification, to account for multi/interdisciplinary research	x is based on (5 year) Impact Factor which has well-documented limitations; x is also vulnerable to scale issues	Using a measure based on citation counts would permit a more meaningful assessment of scientific quality

Table 14.3 (continued)

Indicator	Measure	Computation	Definition	Advantages	Disadvantages	Comment
Hα	Quantity OR quality	The value of hα is equal to N papers with at least αhα citations each and the other $n - H\alpha$ papers have fewer than $\leq \alpha h\alpha$ citations each	Cumulative achievement, advantageous for selected scientists	Greater granularity in comparing scientists with same h; α can be set to the practices in a specific field, allowing for fairer comparison between fields	No agreement on the value of parameter α. The appropriate choice of α requires more study and is field-dependent. Sensitivity of hα to α needs investigating	Small α: ranks scientists based on papers with at least one citation (quantity measure, advantageous for productive scientists who are not highly cited) Large α: measures number of citations of most-cited paper (quality)
Dynamic h	Rank	Builds on three time-dependent elements: $R(T) \times vh(T)$, where R(T) is the R-index computed at time T and vh is the h velocity	Accounts for size and contents of the h-core, number of citations received and the h-velocity	Ideal where two scientists have the same h index and the same number of citations in h-core, where one has no change in h and the other h is on the rise	H-dependent. To define vh, it is better to find a fitting for hrat(t), not for h(t). This function is more similar to a continuous function than the standard h-index	For evaluation purposes, self-citations should be removed
hmx	Rank	Ranks academics by their maximum h (hmx) measured across WoS, Scopus and Google Scholar	Ranking of the academics using all citation databases together	Accounts for missing citations, lack of correlation between databases and disparities in h across databases	Assumes that the differences in h across the databases are due to false-negative errors and that these were negligible	An examination of the overlaps of citations and publications between the databases will provide a better estimate
Gα	Rank AND quality	$g\alpha =$ highest rank the first gα papers have together, at least number of citations	Allows for fractional papers and citations to measure performance at a more precise level	gα-index puts more weight on the quality aspect of scientific performance than the hα-index	No agreement on the value of parameter. The appropriate choice of Gα requires more study and is field-dependent	Empirical research needed to determine whether the gα provides better results than the g-index in practice
g	Rank of scholar	Publications ranked in descending citation order. Cumulative sum of citations is calculated. g is where the square root of the cumulative sum is equal to rank	Cumulative achievement, highly cited papers included, to distinguish between and rank scientists	Corrects h by weighting highly cited papers to make subsequent citations to highly cited papers count in calculation of the index	Can be disproportionate to average publication rate. g of a scientist with one big hit paper and a mediocre core of papers could differ significantly from scientists with a higher average of citations	Ignores the distribution of citations

Table 14.3 (continued)

Indicator	Measure	Computation	Definition	Advantages	Disadvantages	Comment
Rational h	Rank of scholar	hrat = (h + 1) − (nc/2h + 1). h = h-index, nc = number of citations needed to make a h-index of h + 1 and 2	Indicates distance to a higher h by interpolating between h and h + 1. h + 1 = maximum number of cites needed to increase the h index one unit	Increases in smaller steps than h, providing greater distinction in ranking of individuals	Same limitations as for rational g	Interpolated indices have the advantage that one does not have to wait so long to see one's index growing
hg	Rank of scholar	The square root of (the sum of h multiplied by g)	Comparison of scholars with similar h and g indices	Accounts for influence of large successful papers on g-index. Achieves balance between the impact of an author's best papers and the very highly cited ones	Combining h and g does not improve discriminatory power	hg has no direct meaning in terms of papers and citations of a scientist and can lead to hasty judgments

show the same relative or absolute performance, they do not rise or drop in rank position [14.148]. This is not the case with *h*, and already in 2005 *Moed* pointed out that authors with very different citation distributions can have the same *h*-index [14.18], just as many people with very different citation distributions can have the same *CPP*. Consequently, developers continue to suggest new indicators to correct for the inconsistencies in *h*, and use different mathematical manipulations to overcome this. For example, *hw* and *dynamic h* attempt to identify quality papers by applying the Lotkanian approach to concepts of size and rank frequency, aiming to provide steady patterns of the evolution of papers in the *h*-core over time. Thus, no publication can instantly become highly cited, implying that with dynamic indicators, the *h* inconsistency cannot occur [14.161]. However, as can be seen in Table 14.3, many of these hybrid ALIRP are clearly exploratory, in that they raise important questions rather than provide answers for indicator developers. The requirements for data collection and calculation are also more demanding, and this increased complexity does not necessarily correlate with an increase in the validity of the indicator. Further, the focus is not on the practical evaluation of researchers, but on studying the equations within the indicator and the theoretical interpretations of bibliometric laws, stimulating further study [14.160]. These indicators are still experimental, with their own caveats and inconsistencies. As the following examples illustrate, they are not application-ready unless detailed explanations of their shortcomings are provided in evaluation reports. The *h2* [14.166] is an indicator of excellence, but it has a consistency problem, which means that it cannot determine excellence, as it does not discriminate between scientists having different numbers of publications with quite different citation rates for relatively high *h2* indices [14.135]. The *Q2* index [14.167] takes the square root of the sum of the *h*-index multiplied by the *m*-index [14.91] to provide an estimate of the average effect of a set of productive papers. However, *Q2* values are closer to *h* than to *m* (median number of citations to papers in the *h* core), and this can be interpreted as a penalty associated with the *m*-index for researchers where *h* is very low. The *Q2* indicator suffers from the same inconsistency problems as *h* [14.168]. Similarly, the *e*-index [14.169] computes the square root of the surplus of citations in the *h*-set beyond h^2 as a supplement to *h* to indicate excellence. Consequently, *e* only makes sense when *h* is given, which as we know is inherently flawed. Finally, *hg* [14.170] adapts *h* to calculate the square root of the sum of *h* multiplied by the *g*-index to give granular rankings of researchers with similar *h* and *g* indices. Consequently, it includes in its calculation an arbitrary threshold of citations, creating a fractional size–frequency function, and it can be disproportional to the average publication rate [14.3, 171]. This means that the *hg*-index of a scientist with one big hit paper and a mediocre core of papers could increase significantly in comparison to that of scientists with a higher average number of citations. Perhaps the discussion around the validity of ALIRP in this section is too academic, not rooted in real life. Perhaps it is unnecessary to go into so much mathematical detail, and instead to look at the indicator in question and simply consider whether it makes sense. Therefore, taking *CPP* and *%HCP* as examples, they are clearly flawed because of their respective reliance on the arithmetic mean or on WoS disciplinary classifications, but because they are simple to calculate and their failings easily understandable, they are still appropriate for evaluations *if* supplementary indicators and, of course, a narrative are provided. In comparison, the *DCI* index is a highly sophisticated indicator. However, it is so complicated that problems with the indicator are hidden. The difference in *DCI* values can be caused by the weighting parameter, resulting in some authors scoring well on this indicator while others score poorly, with no connection to the currency of their body of work, let alone the quality of their work, which is the concept this indicator is attempting to capture.

The analyses in Tables 14.1–14.3 and discussions in the previous sections show that identifying the appropriate indicator for specific situations is complicated. In bibliometrics, we rely less on anecdotes and favor data, the idea being that numbers tend to lie less than people do when evaluating an individual's academic reputation and career path. However, the demands we place on these numbers must be supported by transparency in the purpose of the evaluation, the robustness of our methods and the validity and clarity of our models, an awareness of what may bias the indicator scores, and openness regarding the limitations of ALIRP. Determining the appropriate application of ALIRP is challenging, and it is the author's hope that Tables 14.1–14.3 will make choosing an ALIRP at least a *little* less complicated.

Assessing the composition and validity of the indicator can help us identify potential structural problems even before application of the indicator in practical evaluation. Coherence in indicator design is a major challenge. One way to improve this situation is to update our understanding of model-building in bibliometrics for both developers of indicators and consumers. This includes the need for clear statements of definition and operationalization of concepts by indicator developers, frank discussion concerning the constraints of indicators, and addressing the technical problems and resource challenges we face in gathering data and computing indicator values in practice. In particular, we as a bibliometric community must address the ad hoc nature of indicator development and provide guidance for the validation and application of indicators to promote coherent evaluation and to prevent conflict. Most importantly, we must strengthen the weak follow-up of indicator development. Not a short list by any means, but this handbook is a practical step in the right direction.

14.5 The Appropriateness of ALIRP and the Application Context

ALIRP are tricky, as they present different bibliometric pictures of researcher performance, and the numerical values demand conscientious interpretation. From a researcher perspective, indicators that are interpreted as portraying them in the most flattering light might be the most useful, but from the evaluator's point of view, indicators that are informative to a particular question are the most useful. This is why the current chapter explores the appropriateness and not the usefulness of ALIRP, as usefulness can differ greatly depending on the end user's needs.

We cannot assume that those responsible for applying an indicator will not be self-serving, manipulating the indicator to appear ineffective or unreliable, or vice versa. The commonality in the application of an indicator, regardless of the end user, is that the interpretation of the ALIRP affects our interpretation of researcher performance. Further, the same indicator can be subject to a variety of interpretations, often based on the many diverse values of the administrators, the culture of the researcher's affiliated organization and the purpose of the evaluation. Therefore, the need for indicators that are valid measures is fundamental to the process of evaluation. But how is validity addressed in the application context? ALIRP are surrogates, measurements taken with the intent to gain insight into something that can be impractical to measure directly or, in principle, impossible to measure, such as a researcher's potential. Impact and influence are common bibliometric measures that have little meaning in and of themselves, but are used to provide insight into the profile of a researcher. With a direct and uncomplicated causal relationship, surrogate measurements are as good as direct measurements. The difficulty arises at the level of the individual, as the relationship between the researcher and the impact, significance or importance of their work is not uncomplicated. Measuring prestige, impact, effect or quality by counting publications and citations is neither a direct nor causal measure, but rather an illusory one. Hence, the need to apply indicators critically and explicitly argues for their validity.

Elkana et al. [14.20], The Acumen Collaboration's portfolio [14.172], and more recently *Belter* [14.173] and *Wilsdon* et al. [14.26], suggest that the way forward is to adopt a balanced approach and discuss a researcher's achievements in terms of spheres of activity, or *domains*. This knowledge is set in a narrative and used to fit appropriate qualitative and quantitative indicators to the situation of the researcher. In the first domain, a manpower and diversity framework for the department in which the researcher is employed is described to set the context and indicate the extent that this environment stimulates the researcher's performance or works against it [14.174]. This includes, among other factors, the size of the department, programmaticity (clearly defined goals and matching strategies, organizational structures, and managerial procedures and abilities that reflect them), the freedom to shift direction and pursue whatever leads appear promising, the hours the researcher must spend teaching, the hours dedicated to research, and perhaps somewhat controversially, the scientific vigor and promise shown by the researcher (documentation of working on weekends and after hours). In the second domain, researchers describe their managerial responsibilities and academic duties, such as reviewing for journals, contributing to internal and external committees, and organizing conferences. The third domain is an explanation of how researchers devote their time to students, what courses they have developed, and supervisory and other responsibilities they have taken on. Fourth, the researchers describe their research work including a description of scientific management (timelines, economy, completion) and the foresight of research, their role in research projects, be it as collaborators, technicians, fund-raisers or project managers, and the publications produced by these projects. *Elkana* et al. [14.20] suggest further fragmentation of this fourth domain, with indicators of how well the individual fits new findings together with previous findings, topic-related indicators, formulation of new questions and solving outstanding questions. Fifth, entrepreneurship and community outreach are described, which can include digital communication of research allowing for the use of web- and altmetrics, taking part in radio broadcasts, invited talks, and holding open house science days for members of the public. Each domain presents the potential for a great variety of indicators and the opportunity to produce rich, contextual information about a researcher's competencies, their innovation and creativity, managerial capabilities, technical skills, methodological strengths, ability to attract funds, social outreach, and economic and pedagogical impact, to name but a few. For each domain, a different palette of indicators is appropriate—including both qualitative and quantitative as well descriptive and evaluative indicators. This evaluation framework puts the role of ALIRP into perspective: note that only one of the aforementioned domains is based on counting scientific publications and citations. ALIRP are thus a small part of a comprehensive program of metrics that work on baseline values of very different value systems, and only together can describe the scholarly and scientific activities of the researcher. This program of indicator application promotes the participation of those affected, and in this way the validity of the chosen metrics is confirmed through collaboration between the evaluator and the evaluand, as recommended in the evaluation principles by *Hicks* et al. [14.25].

14.6 Conclusions

In this chapter, the meaning, methodology and problematic issues associated with ALIRP have been presented. The chapter highlighted the pressing need for monitoring and evaluation of present-day indicator production and improved articulation of the concepts that are operationalized in ALIRP. Constraints and technical problems were exemplified, and the weak follow-up of indicator development was questioned. It was of great value to isolate and deconstruct ALIRP in this chapter, because choosing the appropriate ALIRP is difficult. ALIRP that are valid and suitable for one application may be invalid or of little use in another. If ALRIP are to be used for evaluation, the indicators we choose and how we apply them is important.

Even though no one set of quantitative indicators is in itself sufficient for evaluation or without flaws, that is not to say that such indicators are not necessary or helpful. ALIRP have a role in research evaluation frameworks, as they present a multiplicity of methods, and they increase in value when set in a narrative or with other types of indicators. ALIRP produce numbers, and this numerical information must have context; otherwise, indicators become vulnerable to manipulation and meaningless numerology. The numbers produced by

ALIRP are used as political bricks in a foundation, as a baseline for decision-making. They are considered evidence of a researcher's performance and effect, and as such, their interpretation must not be deduced post hoc. ALIRP are inductive processes that are constantly changing not only with respect to the researcher within the research environment, but also with methods of research administration. Therefore, to be successful, the use of ALIRP must be informed by follow-up research from the bibliometric community, and continued feedback from administrators and from those under evaluation. Successful evaluation with ALIRP is realized in interpersonal contracts between evaluator and evaluand, so trust is essential. It is not our place as bibliometricians to use ALIRP to judge researchers for what they may or may not have done. It is our job to contribute with indicators that can help support meaningful truths in evaluations. ALIRP are only models. They are designed as a numerical representation of something real or something conceptual that is relevant for evaluating a researcher's performance [14.175]. However, as models, they are abstractions, proxies, and they should never be confused with the reality they are trying to model. Thus, if the behavior measured by the ALIRP does not reflect what we see, or if the model does not capture the true aspects of the researcher's performance, it is the model that needs to be fixed—not the researcher.

14.A Appendix

Table 14.4 References to indicators (Tables 14.1–14.3)

Indicator	Bibliographic reference
%HCP	*Costas* et al. (2010) [14.96]
A	*Jin* (2006) [14.176]
a(t)	*Egghe* and *Rousseau* (2000) [14.177]
Adapted pure h	*Chai* et al. (2008) [14.178]
Alternative h	*Batista* et al. (2016) [14.179]
AR	*Jin* et al. (2006) [14.180]
AW	*Harzing* (2016) [14.181]
AWCR	*Harzing* (2016) [14.181]
AWCRpa	*Harzing* (2016) [14.181]
b index	*Brown* (2009) [14.182]
c(t)	*Egghe* and *Rousseau* (2000) [14.177]
Classification of durability	*Costas* et al. (2010) [14.183]
DCI index	*Järvelin* and *Persson* (2008) [14.184]
Dynamic h	*Rousseau* and *Ye* (2009) [14.185]
e	*Zhang* (2009) [14.169]
f	*Tol* (2009) [14.186]
Fc	*Egghe* (2008) [14.187]
g	*Egghe* (2006) [14.188]
Gα	*van Eck* and *Waltman* (2008) [14.189]
h	*Hirsch* (2005) [14.1]
ħ	*Miller* (2006) [14.190]
H index sequences and matrices	*Liang* (2006) [14.191]
H2	*Kosmulski* (2006) [14.166]
Hc	*Sidiropoulos* et al. (2007) [14.192]
hf	*Radicchi* et al. (2008) [14.193]

Table 14.4 (continued)

Indicator	Bibliographic reference
hg	*Alonso* et al. (2010) [14.170]
Hm	*Schreiber* (2008) [14.194]
hmx	*Sanderson* (2008) [14.195]
Hn	*Sidiropoulos* et al. (2007) [14.192]
Hpd	*Kosmulski* (2009) [14.102]
Ht	*Sidiropoulos* et al. (2007) [14.192]
hT index	*Anderson* et al. (2008) [14.162]
hw	*Egghe* and *Rousseau* (2008) [14.196]
Hα	*van Eck* and *Waltman* (2008) [14.189]
Index of age & productivity	*Costas* et al. (2010) [14.96]
IQP	*Antonakis* and *Lalive* (2008) [14.14]
m	*Bornmann* et al. (2008) [14.91]
m-quotient	*Hirsch* (2005) [14.1]
n index	*Namazi* and *Fallahzadeh* (2010) [14.105]
POP h	*Harzing* (2016) [14.181]
Price index	*De Solla Price* (1970) [14.197]
Pure H	*Wan* et al. (2007) [14.198]
Q2	*Cabrerizo* et al. (2010) [14.199]
R	*Jin* et al. (2006) [14.180]
Rational g	*Tol* (2008) [14.200]
Rational h	*Ruane* and *Tol* (2008) [14.201]
t	*Tol* (2009) [14.186]
w-index	*Wu* (2008) [14.202]
x index	*Claro* and *Costa* (2011) [14.104]
π index	*Vinkler* (2009) [14.203]

References

14.1 J.E. Hirsch: An index to quantify an individual's scientific research output, Proc. Natl. Acad. Sci. USA **102**(46), 16569–16572 (2005)

14.2 L. Wildgaard, J.W. Schneider, B. Larsen: A review of the characteristics of 108 author-level bibliometric indicators, Scientometrics **101**(1), 125–158 (2014)

14.3 S. Alonso, F. Cabreriazo, E. Herrera-Viedma, F. Herra: H-index: A review focused in its variants, computation and standardization for different scientific fields, J. Informetr. **3**(4), 273–289 (2009)

14.4 J. Bollen, H. Van de Sompel, A. Hagberg, R. Chute: A principal component analysis of 39 scientific impact measures, PLoS One **4**(6), e6022 (2009)

14.5 F.J. Cabrerizo: H-index and variants, Soft Computing and Intelligent Information Systems, http://sci2s.ugr.es/es/node/36 (2016)

14.6 L. Waltman: A review of the literature on citation impact indicators, J. Informetr. **10**(2), 365–391 (2016)

14.7 A.F.J. van Raan: Measuring science. In: *Handbook of Quantitative Science and Technology Research*, ed. by H. Moed, W. Glänzel, U. Schmoch (Springer, Dordrecht 2004) pp. 19–50

14.8 R. Todeschini, A. Baccini: *Handbook of Bibliometric Indicators: Quantitative Tools for Studying and Evaluating Research* (Wiley-VCH, Weinheim 2016)

14.9 L. Waltman, N.J. van Eck, T.N. van Leeuwen, M.S. Visser, A.F.J. van Raan: Towards a new crown indicator: An empirical analysis, Scientometrics **87**(3), 467–481 (2011)

14.10 D. Hicks: The four literatures of social science. In: *Handbook of Quantitative Science and Technology Research*, ed. by H. Moed, W. Glänzel, U. Schmoch (Springer, Dordrecht 2005) pp. 473–496

14.11 T. Castellani: Epistemological consequences of bibliometrics: Insights from the scientific community, Soc. Epistemol. Rev. Reply Collect. **3**(11), 1–20 (2014)

14.12 M. Grant, A. Booth: A typology of reviews: An analysis of 14 review types and associated methodologies, Health Inf. Libr. J. **26**(2), 91–108 (2009)

14.13 A. Skupin: Discrete and continuous conceptualizations of science: Implications for knowledge domain visualization, J. Informetr. **3**(3), 233–245 (2009)

14.14 J. Antonakis, R. Lalive: Quantifying scholarly impact: IQP versus the Hirsch h, J. Am. Soc. Inf. Sci. Technol. **59**(6), 956–969 (2008)

14.15 R.K. Merton: *The Normative Structure of Science* (University of Chicago Press, Chicago 1973)

14.16 G.N. Gilbert: Referencing as persuasion, Soc. Stud. Sci. **7**(1), 113–122 (1977)

14.17 E. Garfield: *Citation Indexing: Its Theory and Application in Science, Technology and Humanities* (Wiley, New York 1979)

14.18 H.F. Moed: *Citation Analysis in Research Evaluation. Information Science and Knowledge Management* (Springer, Dordrecht 2005)

14.19 A. Singleton: Journal ranking and selection: A review in physics, J. Doc. **32**(4), 258–289 (1976)

14.20 Y. Elkana, J. Lederberg, R. Merton, A. Thackray, H. Zuckerman: *Toward a Metric of Science: The Advent of Science Indicators* (Wiley, New York 1978)

14.21 M.H. MacRoberts, B.R. MacRoberts: Problems of citation analysis, Scientometrics **36**(3), 435–444 (1996)

14.22 I. Podlubny: A note on comparison of scientific impact expressed by the number of citations in different fields of science, Scientometrics **64**(1), 95–99 (2005)

14.23 T. van Leeuwen: Philosophical and theoretical considerations on bibliometric analysis. In: *Proc. NWB2016, Copenhagen* (2016), https://figshare.com/collections/NWB_2016_Oral_Presentations/3581192/2

14.24 L. Leydesdorff, S. Milojevic: Scientometrics, https://arxiv.org/ftp/arxiv/papers/1208/1208.4566.pdf (2012)

14.25 D. Hicks, P. Wouters, L. Waltman, S. de Rijcke, I. Rafols: Bibliometrics: The Leiden Manifesto for research metrics, Nature **520**(7548), 429–431 (2015)

14.26 J. Wilsdon, L. Allen, E. Belfiore, P. Campbell, S. Curry, S. Hill, R. Jones, R. Kain, S. Kerridge, M. Thelwall, J. Tinkler, I. Viney, P. Wouters, J. Hill, B. Johnson: The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management, https://doi.org/10.13140/RG.2.1.4929.1363 (2015)

14.27 A.L. Porter, D.E. Chubin, X.-Y. Jin: Citations and scientific progress: Comparing bibliometric measures with scientist judgments, Scientometrics **13**(3/4), 103–104 (1988)

14.28 A.J. Nederhof, A.F.J. van Raan: Peer review and bibliometric indicators of scientific performance: A comparison of cum laude doctorates with ordinary doctorates in physics, Scientometrics **11**(5), 6–333 (1987)

14.29 H. Moed: A comparative study of bibliometric past performance analysis and peer judgement, Scientometrics **8**(3/4), 149–159 (1985)

14.30 M.E.D. Koenig: Bibliometric indicators versus expert opinion in assessing research performance, J. Am. Soc. Inf. Sci. **34**(2), 136–145 (1983)

14.31 A.J. Nederhof, R.A. Zwaan, R.E. De Bruin, P.J. Dekker: Assessing the usefulness of bibliometric indicators for the humanities and the social and behavioural sciences: A comparative study, Scientometrics **15**(5/6), 423–435 (1989)

14.32 H. Moed: Bibliometric measurement of research performance and Price's theory of differences among the sciences, Scientometrics **15**(5), 473–483 (1989)

14.33 H. Moed: The application of bibliometric indicators: Important field- and time-dependent fac-

tors to be considered, Scientometrics **8**(3), 177–203 (1985)
14.34 E. Garfield: From citation indexes to informetrics: Is the tail now wagging the dog?, Libri **48**(2), 67–80 (1998)
14.35 K.L. Reed: Citation analysis of faculty publication: Beyond science citation index and social science citation index, Bull. Med. Libr. Assoc. **83**(4), 503–508 (1995)
14.36 A. Mendez, I. Gomez, M. Bordons: Some indicators for assessing research performance without citations, Scientometrics **26**(1), 157 (1993)
14.37 X. Xia, M. Li, C.F. Xiao: Author analysis of papers published in "Space Medicine & Medical Engineering" from 1988 to 1998, Space Med. Med. Eng. (Beijing) **12**(6), 431–435 (1999)
14.38 M. Bordons, M.A. Zulueta, A. Cabrero, S. Barrigon: Research performance at the micro level: Analysis of structure and dynamics of pharmacological research teams, Res. Eval. **5**(2), 137–142 (1995)
14.39 H.P.F. Peters, A.F.J. van Raan: A bibliometric profile of top-scientists – a case-study in chemical-engineering, Scientometrics **29**, 115–136 (1994)
14.40 R. Plomp: The highly cited papers of professors as an indicator of a research groups scientific performance, Scientometrics **29**(3), 377–393 (1994)
14.41 P. Banerjee: Indicators of "innovation as process", Scientometrics **43**(3), 331–357 (1998)
14.42 J. Schmoch: Indicators and the relations between science and technology, Scientometrics **38**(1), 103–116 (1997)
14.43 B. Cronin: *The Citation Process: The Role and Significance of Citations in Scientific Communication* (Taylor Graham, London 1984)
14.44 L. Leydesdorff, P. Van den Besselaar: Scientometrics and communication theory: Towards theoretically informed indicators, Scientometrics **38**(1), 155–174 (1997)
14.45 W. Glänzel, U. Schoepflin: Little scientometrics, big scientometrics – and beyond?, Scientometrics **30**(2/3), 375–384 (1994)
14.46 J.S. Katz: Bibliometric standards: Personal experience and lessons learned, Scientometrics **35**(2), 193–197 (1996)
14.47 I.K. Ravichandra Rao: Methodological and conceptual questions of bibliometric standards, Scientometrics **35**(2), 265–270 (1996)
14.48 V.S. Lazarev: On chaos in bibliometric terminology, Scientometrics **35**(2), 271–277 (1996)
14.49 P. Vinkler: Some practical aspects of the standardization of scientometric indicators, Scientometrics **35**(2), 235–245 (1996)
14.50 J.S. Katz, D. Hicks: How much is a collaboration worth? A calibrated bibliometric model, Scientometrics **40**(3), 541–554 (1997)
14.51 B.K. Sen: Mega-authorship from a bibliometric point of view, Malays. J. Libr. Inf. Sci. **2**(2), 9–18 (1997)
14.52 H. Herbertz: Does it pay to cooperate? A bibliometric case study in molecular biology, Scientometrics **33**(1), 117–122 (1995)
14.53 E. Logan: A bibliometric analysis of collaboration in a medical specialty, Scientometrics **20**(3), 417–426 (1991)
14.54 A. Abbasi, J. Altmann, J. Hwang: Evaluating scholars based on their academic collaboration activities: Two indices, the RC-index and the CC-index, for quantifying collaboration activities of researchers and scientific communities, Scientometrics **2010**(83), 1–13 (2014)
14.55 X.Z. Liu, H. Fang: Fairly sharing the credit of multi-authored papers and its application in the modification of h-index and g-index, Scientometrics **91**(1), 37–49 (2012)
14.56 S. Galam: Tailor based allocations for multiple authorship: A fractional gh-index, Scientometrics **89**(1), 365–379 (2011)
14.57 G. Abramo, C.A. D'Angelo, G. Murgia: Variation in research collaboration patterns across academic ranks, Scientometrics **98**(3), 2275–2294 (2014)
14.58 P. Wouters: Beyond the holy grail: From citation theory to indicator theories, Scientometrics **44**(3), 561–580 (1999)
14.59 R. Rousseau: Citation analysis as a theory of friction of polluted air?, Scientometrics **43**(1), 63–67 (1998)
14.60 L. Leydesdorff: Theories of citation?, Scientometrics **43**(1), 5–25 (1998)
14.61 J.M. Russell, R. Rousseau: Bibliometrics and institutional evaluation. In: *Encyclopedia of Life Support Systems (EOLSS) Part 19.3 Science and Technology Policy*, ed. by R. Arvantis (EOLSS, Oxford 2002)
14.62 T.N. van Leeuwen, L.J. van der Wurff, A.F.J. van Raan: The use of combined bibliometric methods in research funding policy, Res. Eval. **10**(3), 195–201 (2001)
14.63 B.D. Cameron: Trends in the usage of ISI bibliometric data: Uses, abuses and implications, Libr. Academy **5**(1), 105–125 (2005)
14.64 R. Costas, M. Bordons: Bibliometric indicators at the micro-level: Some results in the area of natural resources at the Spanish CSIC, Res. Eval. **14**(2), 110–120 (2005)
14.65 Q.L. Burrell: Ambiguity and scientometric measurement: A dissenty view, J. Am. Soc. Inf. Sci. Technol. **52**(12), 1075–1080 (2001)
14.66 D.W. Aksnes, T.B. Olsen, P.O. Seglen: Validation of bibliometric indicators in the field of microbiology: A Norwegian case study, Scientometrics **49**(1), 7–22 (2000)
14.67 K.W. Boyack, K. Börner: Indicator-assisted evaluation and funding of research: visualizing the influence of grants on the number and citation counts of research papers, J. Am. Soc. Inf. Sci. Technol. **54**(5), 447–461 (2003)
14.68 I. Rowlands: Knowledge production, consumption and impact: Policy indicators for a changing world, ASLIB Proceedings **55**(1/2), 5–12 (2003)
14.69 M. Bordons, I. Gomez: One step further in the production of bibliometric indicators at the micro level: Differences by gender and professional

category of scientists, Scientometrics **57**(2), 159–173 (2003)

14.70 S.E. Wiberley Jr.: A methodological approach to developing bibliometric models of types of humanities scholarship, Libr. Q. **73**(2), 121–159 (2003)

14.71 F. Franceshini, D. Maisano, L. Mastrogiacomo: The effect of database dirty data on h-index calculation, Scientometrics **95**(3), 1179–1188 (2013)

14.72 L. Meho, Y. Rogers: Citation counting, citation ranking, and h-index of human-computer interaction researchers: A comparison of Scopus and Web of Science, J. Am. Soc. Inf. Sci. Technol. **59**(11), 1711–1726 (2008)

14.73 A. Noruzi: Google Scholar: The new generation of citation indexes, Libri **55**(4), 170–180 (2005)

14.74 P. Jasco: Google Scholar: The pros and the cons, Online Inf. Rev. **29**(2), 208–214 (2005)

14.75 P. Jasco: As we may search: Comparison of major features of the Web of Science, Scopus, Google Scholar citation-based and citation-enhanced databases, Curr. Sci. **89**(9), 1537–1547 (2005)

14.76 A.W. Harzing: Document categories in the ISI Web of Knowledge: Misunderstanding the social sciences?, Scientometrics **94**(1), 23–34 (2013)

14.77 B. Minasny, A.E. Hartemink, A. McBratney, H.J. Jang: Citations and the h index of soil researchers and journals in the Web of Science, Scopus, Google Scholar, PeerJ **1**, e183 (2013)

14.78 J. Bar-Ilan: Which h-index? – A comparison of WoS, Scopus and Google Scholar, Scientometrics **74**(2), 257–271 (2008)

14.79 T.F. Frandsen, J. Nicolaisen: Intradisciplinary differences in database coverage and the consequences for bibliometric research, J. Am. Soc. Inf. Sci. **59**(10), 1570–1581 (2008)

14.80 M. Schreiber: An empirical investigation of the g-index for 26 physicists in comparison with the h-index, the A-index, and the R-index, J. Am. Soc. Inf. Sci. Technol. **59**(9), 1513–1522 (2008)

14.81 P. Weingart: Impact of bibliometrics on the science system: Inadvertent consequences?, Scientometrics **62**(1), 117–131 (2005)

14.82 H.D. White: Authors as citers over time, J. Am. Soc. Inf. Sci. Technol. **52**(2), 87–108 (2001)

14.83 A.E. Cawkell: Understanding science by analysing its literature, Inf. Sci. **10**(1), 3–10 (1976)

14.84 P. Ball: Index aims for fair ranking of scientists, Nature **436**(7053), 900 (2005)

14.85 T. Marchant: Score-based bibliometric rankings of authors, J. Am. Soc. Inf. Sci. Technol. **60**(6), 1132–1137 (2009)

14.86 J. Panaretos, C.C. Malesios: Assessing scientific research performance and impact with single indices, Scientometrics **81**(3), 635–670 (2014)

14.87 M. Kosmulski: Family-tree of bibliometric indices, J. Informetr. **7**(2), 313–317 (2013)

14.88 M. Schreiber, C.C. Malesios, S. Psarakis: Exploratory factor analysis for the Hirsch index, 17 h-type variants, and some traditional bibliometric indicators, J. Informetr. **6**(3), 347–358 (2012)

14.89 L. Wildgaard: *Measure Up! The Extent Author-Level Bibliometric Indicators are Appropriate Measures of Individual Researcher Performance*, PhD Thesis (Copenhagen University, Copenhagen 2015)

14.90 E.S. Vieira, J.A.S. Cabral, J.A.N.F. Gomes: How good is a model based on bibliometric indicators in predicting the final decisions made by peers?, J. Informetr. **8**(2), 390–405 (2014)

14.91 L. Bornmann, R. Mutz, H. Daniel: Are there better indices for evaluation purposes than the h-index? A comparison of nine different variants of the h-index using data from biomedicine, J. Am. Soc. Inf. Sci. Technol. **59**(5), 830–837 (2008)

14.92 T.N. van Leeuwen: The application of bibliometric analyses in the evaluation of social science research. Who benefits from it, and why it is still feasible, Scientometrics **66**(1), 133–154 (2006)

14.93 P. Juznic, S. Peclin, M. Zaucer, T. Mandelj, M. Pusnik, F. Demsar: Scientometric indicators: Peer-review, bibliometric methods and conflict of interests, Scientometrics **85**(2), 429–441 (2010)

14.94 G. Holden, G. Rosenberg, K. Barker: Bibliometrics, Soc. Work Health Care **41**(3), 4–67 (2005)

14.95 A. Pillay: Academic promotion and the h-index, J. Am. Soc. Inf. Sci. **64**(12), 2598–2599 (2013)

14.96 R. Costas, T.N. van Leeuwen, M. Bordons: A bibliometric classificatory approach for the study and assessment of research performance at the individual level: The effects of age on productivity and impact, J. Am. Soc. Inf. Sci. Technol. **61**(8), 1564–1581 (2010)

14.97 E.S. Vieira, J.A.S. Cabral, J.A.N.F. Gomes: Definition of a model based on bibliometric indicators for assessing applicants to academic positions, J. Am. Soc. Inf. Sci. Technol. **65**(3), 560–577 (2014)

14.98 J.A. Eloy, P. Svider, S.S. Chandrasekhar, Q. Husain, K.M. Mauro, M. Setzen, S. Baredes: Gender disparities in scholarly productivity within academic otolaryngology departments, Otolaryngol. Head Neck Surg. (U.S.) **148**(2), 215–222 (2013)

14.99 U. Sandström, M. Hällsten: Gender; funding diversity and quality of research. In: *Proc. ISSI 2007 – 11th Int. Conf. Int. Soc. Scientometr. Informetr.* (2007) pp. 685–690

14.100 T.F. Frandsen, R.H. Jacobsen, J.A. Wallin, K. Brixen, J. Ousager: Gender differences in scientific performance: A bibliometric matching analysis of Danish health sciences graduates, J. Informetr. **9**(4), 1007–1017 (2015)

14.101 L. Egghe: On the correction of the h-index for career length, Scientometrics **96**(2), 563–571 (2013)

14.102 M. Kosmulski: New seniority-independent Hirsch-type index, J. Informetr. **3**(4), 341–347 (2009)

14.103 A.W. Harzing, S. Alakangas, D. Adams: hIa: An individual annual h-index to accommodate disciplinary and career length differences, Scientometrics **99**(3), 811–821 (2014)

14.104 J. Claro, C.A.V. Costa: A made-to-measure indicator for cross-disciplinary bibliometric ranking of researchers performance, Scientometrics **86**(1), 113–123 (2011)

14.105 M.R. Namazi, M.K. Fallahzadeh: n-Index: A novel and easily-calculable parameter for comparison of researchers working in different scientific fields, Indian J. Dermatol. Venereol. Leprol. **76**(3), 229–230 (2010)
14.106 R. Costas, M. Bordons, T.N. van Leeuwen, A.F.J. van Raan: Scaling rules in the science system: influence of field-specific citation characteristics on the impact of individual researchers, J. Am. Soc. Inf. Sci. Technol. **60**(4), 740–775 (2009)
14.107 D.-Z. Chen, C.-P. Lin, M.-H. Huang, C.Y. Huang: Constructing a new patent bibliometric performance measure by using modified citation rate analyses with dynamic backward citation windows, Scientometrics **82**(1), 149–163 (2014)
14.108 B. Lepori, E. Reale, R. Tijssen: Designing indicators for policy decisions: Challenges, tensions and good practices: Introduction to a special issue, Res. Eval. **20**(1), 3–5 (2011)
14.109 L. Bornmann, R. Mutz: Do we need the h index and its variants in addition to standard bibliometric measures?, J. Am. Soc. Inf. Sci. Technol. **60**(6), 1286–1289 (2009)
14.110 J. Iivari: Expert evaluation vs. bibliometric evaluation: Experiences from Finland, Eur. J. Inf. Syst. **17**, 169–173 (2008)
14.111 W. Glänzel: The need for standards in bibliometric research and technology, Scientometrics **35**(2), 167–176 (1996)
14.112 L. Bornmann, M. Werner: How to evaluate individual researchers working in the natural and life sciences meaningfully? A proposal of methods based on percentiles of citations, Scientometrics **98**(1), 487–509 (2014)
14.113 J.F. Bach: *On the Proper Use of Bibliometrics to Evaluate Individual Researchers* (Académie des Sciences, Paris 2011), http://www.academie-sciences.fr/pdf/rapport/avis170111gb.pdf
14.114 U. Schmoch, T. Schubert, D. Jansen, R. Heidler, R. von Görtz: How to use indicators to measure scientific performance: A balanced approach, Res. Eval. **19**(1), 2–18 (2010)
14.115 E. Sandström, U. Sandström: Meeting the micro-level challenges: Bibliometrics at the individual level. In: *Proc. ISSI 2009 – 12th Int. Conf. Int. Soc. Scientometr. Informetr.*, Vol. 2 (2009) pp. 846–856
14.116 J.L. Ortega: Relationship between altmetric and bibliometric indicators across academic social sites: The case of CSIC's members, J. Informetr. **9**(1), 39–49 (2015)
14.117 A. Bartoli, E. Medveta: Bibliometric evaluation of researchers in the internet age, Inf. Soc. **30**(5), 349–354 (2014)
14.118 H. Browman, K. Stergiou: Factors and indices are one thing, deciding who is scholarly, why they are scholarly, and the relative value of their scholarship is something else entirely, Ethics Sci. Env. Politics **8**, 1–3 (2014)
14.119 T.N. van Leeuwen: Testing the validity of the Hirsch-index for research assessment purposes, Res. Eval. **17**(2), 157–160 (2014)
14.120 R. Costas, T.N. van Leeuwen, M. Bordons: Self-citations at the meso and individual levels: Effects of different calculation methods, Scientometrics **82**(3), 517–537 (2010)
14.121 M. Schreiber: Examples for counterintuitive behavior of the new citation-rank indicator P100 for bibliometric evaluations, J. Informetr. **8**(3), 738–748 (2014)
14.122 N. Gaster, M. Gaster: A critical assessment of the h-index, BioEssays **34**(10), 830–832 (2012)
14.123 C. Michels, U. Schmoch: Impact of bibliometric studies on the publication behaviour of authors, Scientometrics **98**(1), 369–385 (2014)
14.124 H.P.W. Bauer, G. Schui, A. von Eye, G. Krampen: How does scientific success relate to individual and organizational characteristics? A scientometric study of psychology researchers in the German-speaking countries, Scientometrics **94**(2), 523–539 (2013)
14.125 C. Bloch, J.W. Schneider: Performance-based funding models and researcher behavior: An analysis of the influence of the Norwegian Publication Indicator at the individual level, Res. Eval. **25**(4), 371–382 (2016)
14.126 A. Ebadi, A. Schiffauerova: How to become an important player in scientific collaboration networks?, J. Informetr. **9**(4), 809–825 (2015)
14.127 A. Dilger, L. Lütkenhöner, H. Müller: Scholars' physical appearance, research performance, and feelings of happiness, Scientometrics **104**(2), 555–573 (2015)
14.128 D.R. Amancio, O.N. Oliveira Jr., L. da Fontoura Costa: Topological-collaborative approach for disambiguating authors' names in collaborative networks, Scientometrics **102**(1), 465–485 (2015)
14.129 M. Bordons, J. Aparicio, B. González-Albo, A.A. Díaz-Faes: The relationship between the research performance of scientists and their position in co-authorship networks in three fields, J. Informetr. **9**(1), 135–144 (2015)
14.130 J. Kim, J. Kim: Rethinking the comparison of coauthorship credit allocation schemes, J. Informetr. **9**(3), 667–673 (2015)
14.131 J.C. Nabout, M.R. Parreira, F.B. Teresa, F.M. Carneiro, H.F. da Cunha, L. de Souza Ondei, S.S. Caramori, T.N. Soares: Publish (in a group) or perish (alone): The trend from single- to multi-authorship in biological papers, Scientometrics **102**(1), 357–364 (2015)
14.132 A.W. Harzing: Health warning: Might contain multiple personalities—the problem of homonyms in Thomson Reuters Essential Science Indicators, Scientometrics **105**(3), 2259–2270 (2015)
14.133 J.W. Schneider: Caveats for using statistical significance tests in research assessments, J. Informetr. **7**(1), 50–62 (2013)
14.134 J.W. Schneider: Null hypothesis significance tests. A mix-up of two different theories: The basis for widespread confusion and numerous misinterpretations, Scientometrics **102**(1), 411–432 (2015)

14.135 L. Waltman, N.J. van Eck: The inconsistency of the h-index, J. Am. Soc. Inf. Sci. Technol. **63**(2), 406–415 (2012)
14.136 G. Prathap: The inconsistency of the h-index, J. Am. Soc. Inf. Sci. Technol. **63**(7), 1480–1481 (2012)
14.137 L. Bornmann, H.D. Daniel: Convergent validation of peer review decisions using the h index: Extent of and reasons for type I and type II errors, J. Informetr. **1**(3), 204–213 (2007)
14.138 A.F.J. van Raan: The influence of international collaboration on the impact of research results, Scientometrics **42**(3), 423–428 (1998)
14.139 W. Glänzel, A. Schubert: A characterization of scientometric distributions based on harmonic means, Scientometrics **26**(1), 81–96 (1992)
14.140 J.W. Schneider, T.N. van Leeuwen: Analysing robustness and uncertainty levels of bibliometric performance statistics supporting science policy. A case study evaluating Danish postdoctoral funding, Res. Eval. **23**(4), 285–297 (2014)
14.141 W. Glänzel: High-end performance or outlier? Evaluating the tail of scientometric distributions, Scientometrics **97**(1), 13–23 (2013)
14.142 M. Schreiber: How much do different ways of calculating percentiles influence the derived performance indicators? A case study, Scientometrics **97**(3), 821–829 (2013)
14.143 D. Bouyssou, T. Marchant: An axiomatic approach to bibliometric rankings and indices, J. Informetr. **8**(3), 449–477 (2014)
14.144 G. Prathap: Quantity, quality, and consistency as bibliometric indicators, J. Am. Soc. Inf. Sci. Technol. **65**(1), 214 (2014)
14.145 L. Egghe, R. Rousseau: A basic indicator for relative contributions and a remarkable difference between a ratio of averages and an average of ratios, Malays. J. Libr. Inf. Sci. **19**(1), 17–22 (2014)
14.146 J. Schulz: Using Monte Carlo simulations to assess the impact of author name disambiguation quality on different bibliometric analyses, Scientometrics **107**(3), 1283–1298 (2016)
14.147 W. Marx, L. Bornmann: On the problems of dealing with bibliometric data, J. Am. Soc. Inf. Sci. Technol. **65**(4), 866–867 (2014)
14.148 E. Riviera: *Mapping scientific literature: structuring scientific communities through Scientometrics*, Ph.D. Thesis (Università degli Studi di Milano Biocca, Milano 2012)
14.149 E. Delgado López-Cózar, N. Robinson-García, D. Torres-Salinas: The Google scholar experiment: How to index false papers and manipulate bibliometric indicators, J. Am. Soc. Inf. Sci. Technol. **65**(3), 446–454 (2014)
14.150 Z. Taşkın, U. Al: Standardization problem of author affiliations in citation indexes, Scientometrics **98**(1), 347–368 (2014)
14.151 L. Wildgaard: A critical cluster analysis of 44 indicators of author-level performance, J. Informetr. **10**(4), 1055–1078 (2016)
14.152 L. Wildgaard: A comparison of 17 author-level bibliometric indicators for researchers in astronomy, environmental science, philosophy and public health in Web of Science and Google Scholar, Scientometrics **104**(3), 873–906 (2015)
14.153 A.F.J. van Raan: Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment of 147 chemistry research groups, Scientometrics **67**(3), 491–502 (2006)
14.154 W. Glänzel, H.F. Moed: Opinion paper: Thoughts and facts on bibliometric indicators, Scientometrics **96**(1), 381–394 (2013)
14.155 J.K. Vanclay: On the robustness of the h-index, J. Am. Soc. Inf. Sci. Technol. **58**(10), 1547–1550 (2015)
14.156 Merriam Webster Dictionary: "model", https://www.merriam-webster.com/dictionary/model
14.157 Y. Gingras: Criteria for evaluating indicators. In: *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*, ed. by B. Cronin, C. Sugimoto (MIT Press, Cambridge 2014) pp. 109–125
14.158 L. Bornmann: How to analyze percentile citation impact data meaningfully in bibliometrics: The statistical analysis of distributions, percentile rank classes, and top-cited papers, J. Am. Soc. Inf. Sci. Technol. **64**(3), 587–595 (2013)
14.159 P. Albarran, I. Ortuno, J. Ruiz-Castillo: Average-based versus high- and low-impact indicators for the evaluation of scientific distributions, Res. Eval. **20**(4), 325–339 (2011)
14.160 F. Ye: A Theoretical approach to the unification of informetric models by wave-heat equations, J. Am. Soc. Inf. Sci. Technol. **62**(6), 1208–1211 (2011)
14.161 F. Ye: H-inconsistency is not an issue in dynamical systems, ISSI Newsletter **8**(2), 22–24 (2012)
14.162 T.R. Anderson, R.K.S. Hankin, P.D. Killworth: Beyond the durfee square: Enhancing the h-index to score total publication output, Scientometrics **76**(3), 577–588 (2008)
14.163 G.J.J. Biesta: *Good Education in an Age of Measurement: Ethics, Politics, Democracy* (Routledge, New York 2010)
14.164 R. Clarke, A. Pucihar: The Web of Science revisited: Is it a tenable source for the information systems discipline or for eCommerce researchers?, http://www.rogerclarke.com/SOS/WoSRev.html (2012)
14.165 IQP Calculator: http://www.hec.unil.ch/jantonakis/IQp%20calculator%20version%202008.xls
14.166 M. Kosmulski: A new type Hirsch-index saves time and works equally well as the original h-index, ISSI Newsletter **2**(3), 4–6 (2006)
14.167 F.J. Caberizo, S. Alonso, E. Herrera-Viedma, F. Herrera: Q2-index: Quantitative and qualitative evolution based on the number and impact of papers in the Hirsch core, J. Informetr. **4**, 23–28 (2012)
14.168 A. Rubem, A. de Moura, J. Soares de Mello: Comparative analysis of some individual bibliometric indices when applied to groups of researchers, Scientometrics **102**(1), 1091–1035 (2015)
14.169 C.-T. Zhang: The e-index, complementing the h-index for excess citations, PLoS ONE **4**(5), e5429 (2009)

14.170 S. Alonso, F.J. Caberizo, E. Herrera-Viedma, F. Herrera: Hg-index: A new index to characterize the scientific output of researchers based on the h- and g-indices, Scientometrics **82**, 391–400 (2010)
14.171 R. Costas, M. Bordons: The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level, J. Informetr. **1**(3), 193–203 (2007)
14.172 Academic careers understood through measurements and norms (ACUMEN): Portfolio, http://research-acumen.eu/portfolio
14.173 C.W. Belter: Bibliometric indicators: Opportunities and limits, J. Med. Libr. Assoc. **103**(4), 219–221 (2015)
14.174 R.K. Swihart, M. Sundaram, T.O. Höök, J.A. Dewoody: Factors affecting scholarly performance by wildlife and fisheries faculty: Factors affecting bibliometrics, J. Wildl. Manag. **80**(3), 563–572 (2016)
14.175 C. Dym: *Principles of Mathematical Modelling*, 2nd edn. (Academic, New York 2004)
14.176 B.H. Jin: H-index: An evaluation indicator proposed by scientist, Sci. Focus **1**(1), 8–9 (2006)
14.177 L. Egghe, R. Rousseau: Aging, obsolescence, impact, growth, and utilization: Definitions and relations, J. Am. Soc. Inf. Sci. Technol. **51**(11), 1004–1017 (2000)
14.178 J. Chai, P. Hua, R. Rousseau, J. Wan: The adapted pure h-index. In: *Proc. WIS 2008: 4th Int. Conf. Webmetrics, Informetr. Scientometr. 9th COLLNET Meeting, Berlin* (2008), http://www.collnet.de/Berlin-2008/ChaiWIS2008aph.pdf
14.179 P. Batista, M. Campiteli, O. Kinouchi, A. Martinez: Is it possible to compare researchers with different scientific interests?, Scientometrics **68**(1), 179–189 (2006)
14.180 B.H. Jin, L.L. Liang, R. Rousseau, L. Egghe: The R and AR indices: Complementing the h-index, Chin. Sci. Bull. **52**(6), 855–863 (2007)
14.181 A.W. Harzing: Publish or Perish user's manual, http://www.harzing.com/pophelp/metrics.htm (2016)
14.182 R. Brown: A simple method for excluding self-citations from the h-index: The b-index, Online Inf. Rev. **33**(6), 1129–1136 (2009)
14.183 R. Costas, T.N. van Leeuwen, A.F.J. van Raan: Is scientific literature subject to a sell by date? A general methodology to analyze the durability of scientific documents, J. Am. Soc. Inf. Sci. Technol. **61**(2), 329–339 (2010)
14.184 K. Järvelin, O. Persson: The DCI-index: Discounted cumulated impact based on research evaluation, J. Am. Soc. Inf. Sci. Technol. **59**(9), 1433–1440 (2008)
14.185 R. Rousseau, F.Y. Ye: A proposal for a dynamic h-type index, J. Am. Soc. Inf. Sci. Technol. **59**(11), 1853–1855 (2008)
14.186 R.S.J. Tol: Of the H-index and its alternatives: An application to the 100 most prolific economists, Scientometrics **80**(2), 317–324 (2009)
14.187 L. Egghe: Mathematical theory of the h- and g-index in case of fractional counting of authorship, J. Am. Soc. Inf. Sci. Technol. **59**(10), 1608–1616 (2008)
14.188 L. Egghe: Theory and practise of the g-index, Scientometrics **69**(1), 131–152 (2006)
14.189 N.J. van Eck, L. Waltman: Generalizing the g- and h-indices, J. Informetr. **2**(4), 263–271 (2008)
14.190 C.W. Miller: Superiority of the h-index over the impact factor for physics, arXiv:physics/0608183 [physics.soc-ph] (2006)
14.191 L. Liang: H-index sequence and h-index matrix: Constructions and applications, Scientometrics **69**(1), 153–159 (2006)
14.192 A. Sidiropoulos, D. Katsaros, Y. Manolopoulos: Generalized hirsch h-index for disclosing latent facts in citation networks, Scientometrics **72**(2), 253–280 (2007)
14.193 F. Radicchi, S. Fortunato, C. Castellano: Universality of citation distributions: Toward an objective measure of scientific impact, Proc. Natl. Acad. Sci. USA **105**(45), 17268–17272 (2008)
14.194 M. Schreiber: A modification of the h-index: The H(m)-index accounts for multi-authored manuscripts, arXiv:0805.2000 [Physics.Soc-Ph] (2008)
14.195 M. Sanderson: Revisiting h measured on UK LIS and IR academics, J. Am. Soc. Inf. Sci. Technol. **59**(7), 1184–1190 (2008)
14.196 L. Egghe, R. Rousseau: An h-index weighted by citation impact, Inf. Process. Manag. **44**(2), 770–780 (2008)
14.197 D.J.D. Price: Citation measures of hard science, soft science, technology and non-science. In: *Communication Among Scientists and Engineers*, ed. by C.E. Nelson, D.K. Pollack (Heath Lexington, Lexington 1970) pp. 3–22
14.198 J. Wan, P. Hua, R. Rousseau: The pure h-index: Calculating an author's h- index by taking co-authors into account, COLLNET J. Scientometr. Inf. Manag. **1**(2), 1–5 (2007)
14.199 F.J. Cabrerizo, S. Alonso, E. Herrera-Viedmac, F. Herrera: Q2-index: Quantitative and qualitative evaluation based on the number and impact of papers in the hirsch core, J. Informetr. **4**(1), 23–28 (2010)
14.200 R.S.J. Tol: A rational, successive g-index applied to economics departments in Ireland, J. Informetr. **2**(2), 149–155 (2008)
14.201 F. Ruane, R. Tol: Rational (successive) h-indices: An application to economics in the Republic of Ireland, Scientometrics **75**(2), 395–405 (2008)
14.202 Q. Wu: The w-index: A significant improvement of the h-index, arXiv:0805.4650v1 [Physics.Soc-Ph] (2008)
14.203 P. Vinkler: The π-index: A new indicator for assessing scientific impact, J. Inf. Sci. **35**(5), 602–612 (2009)

Lorna Wildgaard

Researcher Support
The Royal Library
Copenhagen, Denmark
lowi@kb.dk

Lorna Wildgaard is a Data Management Specialist at the Royal Library, researcher services at Copenhagen University. She obtained her PhD in bibliometrics in 2015. Her research areas include bibliometrics andresearch evaluation, medical informatics, research methods, information seeking, communication of information through visualmedia, and the potential of eyetracking in understanding the communication of information. She is passionate about teaching, co-organizer of DST4L2016, and co-founder of a new data science network in Denmark for librarians.

15. Challenges, Approaches and Solutions in Data Integration for Research and Innovation

Maurizio Lenzerini, Cinzia Daraio

In order to be implemented by policy makers, science, technology, and innovation (STI) policies and indicator building need data. Whenever we need data, we need a method for data management, and in the era of big data, a crucial role is played by data integration. Therefore, STI policies and indicator development need data integration. Two main approaches to data integration exist, namely procedural and declarative. In this chapter, we follow the latter approach and focus our attention on the ontology-based data integration (OBDI) paradigm. The main principles of OBDI are:

(i) Leave the data where they are.
(ii) Build a conceptual specification of the domain of interest (ontology), in terms of knowledge structures.
(iii) Map such knowledge structures to concrete data sources.
(iv) Express all services over the abstract representation.
(v) Automatically translate knowledge services to data services.

We introduce the main challenges of data integration for research and innovation (R&I) and show that reasoning over an ontology connected to data may be very helpful for the study of R&I. We also provide examples by using *Sapientia*, an ontology specifically defined for multidimensional research assessment.

15.1 The Role of Data Integration for Research and Innovation

In the last years, the amount of data available for research and innovation (R&I) is growing, in particular thanks to data collections and other initiatives of international and national organizations. While the availability of data stored in current information systems and the processes making use of such data are exponentially increasing, turning this data into information and governing both data and processes are still great challenges in the context of information technology (IT) [15.1–3]. These issues arise from the proliferation of data sources and services both within a single organization and in cooperating environments. Data integration and data interoperability, which have been important in the last decades, are even more important today, in the big data era (see, for instance, two recent books on big data integration [15.4, 5]). Some of the theoretical issues that are relevant for data integration and data interoperability are modeling a data integration application, extract-

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_15

ing and exchanging data from relevant sources, dealing with inconsistent data sources, and processing and reasoning on queries [15.6].

According to *Parent* and *Spaccapietra* [15.7], interoperability is the way in which heterogeneous systems talk to each other and exchange information in a meaningful way. They recognized three stages of interoperability, from the lowest based on no integration, to an intermediary stage in which the system does not guarantee consistency across database boundaries, to a higher stage, which has the objective of developing a global system embracing the existing systems, in order to deliver the desired level of integration of the data sources.

Two main approaches to data integration exist, namely procedural and declarative. In the procedural approach, also called bottom-up approach, for every *information need*, one figures out which data are needed and how they can be accessed, and the goal is to design and realize the corresponding service. On the other hand, in the declarative approach, also called top-down approach, one defines a global representation structure that is valid for the domain of interest underlying the data sources, links this structure to the actual data, lets the user use this structure to specify the *information needs*, and the goal is to automatically extract the right data from the sources. In this chapter, we follow the latter approach and focus our attention to the ontology-based data integration (OBDI) paradigm, which is a recently introduced declarative paradigm for data integration and governance. OBDI uses knowledge representation and reasoning techniques for a new way of integrating and governing data. The principles at the basis of OBDI can be summarized as follows:

(i) Leave the data where they are.
(ii) Build a conceptual specification of the domain of interest, in terms of knowledge structures; such a conceptual representation is called ontology.
(iii) Map such knowledge structures to concrete data sources.
(iv) Express all services over the abstract representation.
(v) Automatically translate knowledge services to data services.

An OBDI system is thus constituted by three main components: the ontology, which represents a conceptual description of the domain; the data sources, where the actual data are, and the mappings that link the data sources to the ontology. Additionally, the ontology is expressed in a form that is both computational and logical. The computational form allows the ontology not only to be understood by humans, but also to be manipulated by the computer, to aid human and machine agents in their performance of tasks within that domain. The logical form is instrumental in enabling additional properties to be inferred by logical reasoning. More generally, reasoning can be used for different goals, such as verification, validation, analysis, synthesis, and exploitation of the latest development in automated reasoning.

The main benefits of the OBDI approach for R&I, as we will see in the examples reported in the following, are related to the opportunity of reasoning over the conceptual structure of the domain (the ontology), reasoning over the mappings of the data sources to the ontology, and reasoning over the data and indicators, for their consistency analysis, their validation, and their data quality assessment (see also the conclusions for an extended summary).

Data integration for R&I is a challenging issue because R&I activities are complex and their assessment is complex too. This is because it requires a systemic approach in which research activities are considered together with education and innovation activities. Moreover, the development of models of indicators or metrics requires a comprehensive framework that includes the specification of the underlying theory, methodology, and data dimensions. Models of metrics are necessary to assess the meaning, validity, and robustness of metrics [15.8]. The complexity of R&I assessment also arises from the consideration of the *implementation* problem according to this three-dimensional framework (see [15.8] for more details).

A workshop organized during the *15th International Conference on Scientometrics and Informetrics* held in Istanbul (Turkey) on 29 June–4 July 2015 discussed the Grand Challenges in Data Integration for Research and Innovation (R&I) Policy. The grand challenges identified were: handling big data, coping with quality issues and anticipating new policy needs.

The analysis of data integration for R&I policy was framed on a groundwork scheme composed by four main areas of intervention and a list of critical issues [15.9]. The main four areas of intervention are:

- Data collection/project initiatives
- Open data, linked data, and platforms for STI
- Monitoring performance evaluation
- Stakeholders, actions, options, costs and sustainability.

The identified critical issues, without being fully comprehensive, are:

- Data quality (considered as *fitness for use* with respect to user needs, see [15.10]) issues – completeness, validity, accuracy, consistency, availability and timeliness

- Comparability problems related to heterogeneous definitions of the variables, data collection practices and databases
- Lack of standardization
- Lack of interoperability
- Lack of modularization
- Problems of classification
- Difficulties in the creation of concordance tables among different classification schemes
- Problems and costs of the extensibility of the system
- Problems and costs of the updating of the system.

Interestingly, many of these issues were already discussed in a Special Session of the ISSI Conference in 1995 in Chicago [15.11]. Moreover, the need for harmonization and standardization of data in R&I is discussed in *Glänzel* and *Willems* [15.12]. It seems that the need for "a clear and unambiguous terminology and specific standards" [15.13, p. 176] is still relevant and timely nowadays.

As described in *Daraio* and *Glänzel* [15.9], the complexity of R&I systems requires a continuous information exchange. This process is due to the communication and interaction process among all actors and agencies involved in the production, processing, and application of knowledge. All data entries, all processing, development, and application of data relevant for research, technology, and innovation have their own rules and standards.

Figure 15.1 shows some elementary rules of interferences expressed in terms of data definition and standard setting in the process of data integration for different purposes, including process monitoring, input–output monitoring, and ex-ante and ex-post evaluation. The application of appropriate standards and data harmonization is crucial to achieve the interoperability of heterogeneous sources of data.

The chapter unfolds as follows. The next section presents the problem of data integration and data governance in a general way. Section 3 describes a formal framework for ontology-based data integration. Section 4 presents the ontology *Sapientia* and the OBDI system developed for multidimensional research assessment. Section 5 presents some illustrative examples of reasoning over *Sapientia*, while Section 6 summarizes the main points and concludes the chapter.

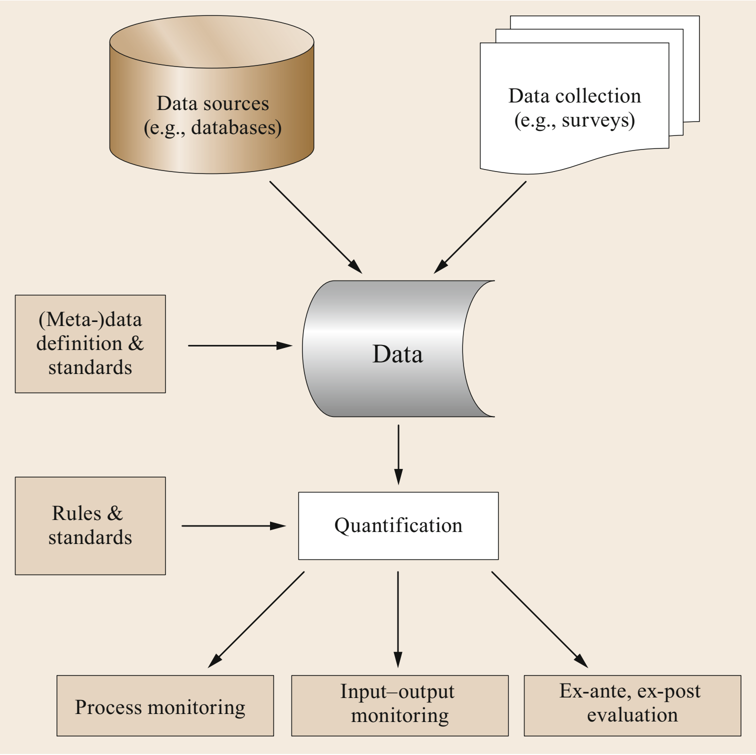


Fig. 15.1 Data integration in use for different purposes with interference points for standardization (after *Daraio* and *Glänzel* [15.9])

15.2 The Problem of Data Integration and Data Governance

Big data management and analysis form a key technology for the competitive advantage of today's enterprises and for shaping the future data-driven society. However, after years of focus on technologies for big data storing and processing, many observers are pointing out that making sense of big data cannot be done without suitable tools for conceptualizing, repairing, and integrating data (http://www.dbta.com/). A common opinion in technology observers is that big data are ready for analysis; one should simply access, select, and load data from big data sources, and magically gain insight, discover patterns, and extract useful knowledge. As pointed out in [15.14], this is not the case; loading a big data platform with quality data with enough structure to deliver value is a lot of work. Thus, it is not surprising that data scientists spend a comparatively large amount of time in the data preparation phase of a project. Whether you call it data wrangling, data preparation, or data integration, it is estimated that 50–80% of a data scientists' time is spent on preparing data for analysis. If we consider that in any IT organization, data governance is also essential for tasks other than data analytics, we can conclude that the challenge of identifying, collecting, retaining, and providing access to all relevant data for the business at an acceptable cost, is huge. If we specialize the above observation to the domain of interest in this chapter, namely R&I, we can state that one of the most challenging tasks for carrying out quantitative studies in the realm of R&I is to identify, collect, integrate, organize, govern and access all relevant data.

Although data integration is one of the oldest problems in data management, the above observations show that it is a major challenge today. As we said in Sect. 15.1, in principle, there are two main approaches to this problem: procedural and declarative. In the procedural approach, sometimes called bottom-up approach, whenever an information need arises that requires accessing the integrated data, a specific program is coded, so as its execution produces the required information. In some sense, with this approach, integration is achieved on a query-by-query basis. In the declarative approach (top-down), one defines a priori an integration database structure, and in order to satisfy an information need one can simply pose a query over such a structure (Fig. 15.2 illustrates these notions relative to R&I). So, with this approach, integration is achieved independently from a specific query (or a specific set of queries).

In this chapter, we focus on the declarative approach and we refer to the typical architecture underlying this approach, which is based on three components [15.6, 16]: the global schema, the sources, and the mapping between the two. The sources represent the repositories

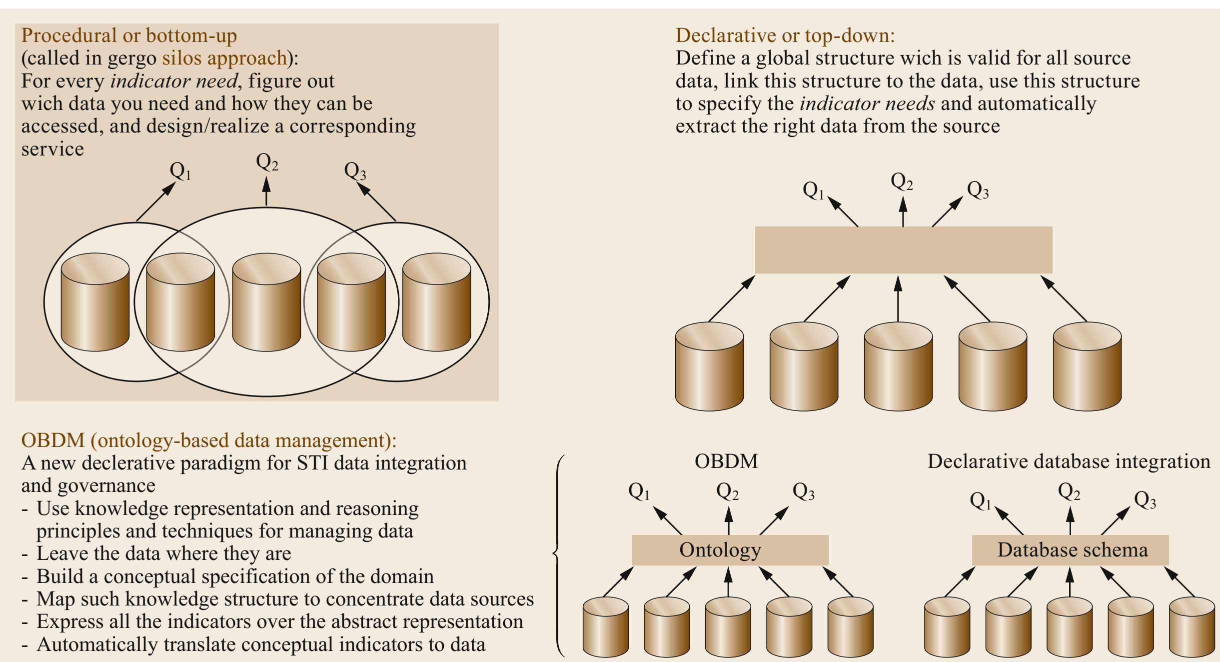


Fig. 15.2 Approaches to data integration for STI (after [15.15])

where the data are; the global schema, also called mediated schema, represents the unified structure presented to the clients; the mapping relates the source data with the global schema. The most important service provided by the integration system is query answering, i. e., computing the answer to a query posed in terms of the global schema. Such computation involves accessing the sources, collecting the relevant data, and packaging such data in the final answer.

Formal, declarative approaches to data integration started in the 1990s [15.6, 16–18]. Since then, many aspects of the general problems have been the subject of detailed investigations both in academia and in industry.

Among them, in this chapter, we want to focus on the idea of using semantics for making data integration more powerful. As illustrated in [15.14], using semantics here means conceiving data integration systems where the semantics of data is explicitly specified and is taken into account to devise all the functionalities of the system. Over the past two decades, this idea has become increasingly crucial to a wide variety of information-processing applications and has received much attention in the artificial intelligence, database, web, and data mining communities [15.19].

As we said before, we concentrate on a specific paradigm for semantic data integration, OBDI. This new paradigm was introduced about a decade ago, as a new way for modeling and interacting with a data integration system [15.20–23]. According to such paradigm, the client of the information system is freed from being aware of how data and processes are structured in concrete resources (databases, software programs, services, etc.) and interacts with the system by expressing her queries and goals in terms of a conceptual representation of the domain of interest, called ontology. An OBDI system is an information management system maintained and used by a given organization (or a community of users), whose architecture has the same structure as a typical data integration system, with the following components: an ontology, a set of data sources, and the mapping between the two. The ontology approach as outlined in the following paragraphs was originally published in [15.24].

The ontology is a conceptual, formal description of the domain of interest to the organization, expressed in terms of relevant concepts, attributes of concepts, relationships between concepts, and logical assertions formally describing the domain knowledge. The data sources are the repositories accessible by the organization where data concerning the domain are stored. In the general case, such repositories are numerous and heterogeneous, each one managed and maintained independently from the others. It may be even the case that some of the data sources are not under the control of the organization and can be accessed remotely, e. g., via the web. The mapping is a precise specification of the correspondence between the data contained in the data sources and the elements of the ontology, where by element we here mean concepts, attributes, and relationships. The main purpose of an OBDI system is to allow information consumers to query the data using the elements in the ontology as predicates. In the special case where the organization manages a single data source, the term *ontology-based data access* (OBDA) system is used instead of the OBDI system.

The notions of OBDA and OBDI were introduced in [15.20–22] and originated from several disciplines, in particular, information integration, knowledge representation and reasoning, and incomplete and deductive databases. OBDI can be seen as a sophisticated form of information integration, where the usual global schema is replaced by an ontology describing the domain of interest. The main difference between OBDI and traditional data integration is that in the OBDI approach, the integrated view that the system provides to information consumers is not merely a data structure accommodating the various data at the sources, but a semantically rich description of the relevant concepts in the domain of interest, as well as the relationships between such concepts. In general, such a description is formally defined in logic and enriches, generalizes, and relates the vocabularies of different data sources, thus providing a *common ground* for the domain knowledge. Also, the distinction between the ontology and the data sources reflects the separation between the conceptual level, the one presented to the client, and the logical/physical level of the information system, the one stored in the sources, with the mapping acting as the reconciling structure between the two levels. Notably, this separation is also instrumental for recovering the possibility of access to legacy systems, which are often excluded from the possibility of being used in interesting analyses carried out by the organization on its own data.

From all the above observations one can easily see that the central notion of OBDI is the ontology, and, therefore, reasoning over the ontology is at the basis of all the tasks that an OBDI system has to carry out. By reasoning, we mean the ability to derive all the implicit knowledge on the domain that is logically implied by the explicitly asserted facts in the ontology. In particular, the axioms of the ontology allow one to derive new facts from the source data, and these inferred facts greatly influence the set of answers that the system should compute during query processing. In the last decades, research on ontology languages and ontology inferencing has been very active in the area of knowledge representation and reasoning. Description logics (DLs) [15.25] are widely recognized as appropriate logics for expressing ontologies and are at the basis of the W3C standard ontology language OWL. These logics permit the specification of a domain by providing the definition of classes and by structuring the knowledge about the classes using a rich set of logical operators. They are decidable fragments of mathematical logic, resulting from extensive investigations on the trade-off between expressive power of knowledge representation languages and computational complexity of reasoning tasks. Indeed, the constructs appearing in the DLs used in OBDI have been carefully chosen taking into account such a trade-off.

As we said before, the axioms in the ontology can be seen as semantic rules that are used to complete the knowledge given by the raw facts determined by the data in the sources. In this sense, the source data of an OBDI system can be seen as an incomplete database, and query answering can be seen as the process of computing the answers logically deriving from the combination of such incomplete knowledge and the ontology axioms. Therefore, at least conceptually, there is a connection between OBDI and the two areas of incomplete information [15.26] and deductive databases [15.27]. The new aspect of OBDI is related to the kind of incomplete knowledge represented in the ontology, which differs both from the formalisms typically used in databases under incomplete information (e. g., Codd tables) and from the rules expressible in deductive database languages (e. g., logic programming rules).

Once we are able to reason about the ontology, we can take advantage of the OBDI system in many interesting and relevant ways. Some of these are referred to below:

1. As already noticed, we can take into account all inferences over the ontology in processing the queries. In other words, the quality of the query answering service is potentially much higher than in the traditional setting, because all the knowledge about the domain represented by the ontology is exploited in computing the answers.
2. We can carry out the task of data source profiling again by relying on the knowledge about the domain. This allows the data designer to describe, maintain, and document the content of the various data sources in a much richer way than in traditional systems, because (s)he can now specify the characteristics of the sources in terms of the vocabulary and the metadata sanctioned by the ontology.
3. We can check and assess the quality of the data sources by comparing their content and their structure with the ontology, and, therefore, singling out inconsistencies, incompleteness, and inaccuracies of the data sources with respect to the domain knowledge.
4. We can set up new interesting services realized through the integration system. One notable example is open data publishing. The presence of the ontology makes it simple and effective to annotate the published data with the concepts and the relationships that are relevant in the domain of interest, so as to provide a conceptual description and a meaningful context of the published datasets.

15.3 Formal Framework for OBDI

The previous section introduced the notion of OBDI as a new paradigm for data integration. In this section, we provide the fundamental elements for a formalization of OBDI, illustrating the form of an OBDI specification, and presenting the semantics of an OBDI system. We closely follow the exposition in [15.14]. In the formal framework, we assume that the OBDI system can access the data sources through a single SQL (Structured Query Language) interface, which presents the various data as if they were in a unique database. In other words, we talk about a single data source, which is a database obtained as an abstraction for a variety of (possibly heterogeneous) data sources. This is not a real limitation, because in practice, such a database might be obtained through the use of an off-the-shelf data federation tool, which presents the sources through a schema of a single database as a wrapping of the source schemas expressed in terms of a unique format.

Given this assumption, an *OBDI specification* I is as a triple $\langle O, S, M \rangle$, where O is an ontology, S is a relational schema, called source schema, and M is a mapping from S to O. As already stated, O represents the general knowledge about the domain (i. e., relative to classes and relationships, rather than to specific objects that are instances of concepts), expressed in some logical language. Typically, O is a lightweight DL TBox [15.20], i. e., it is expressed in a language ensuring both semantic richness and efficiency of reasoning, and in particular of query answering. The mapping

SubClassOf(*Dean Professor*)
SubClassOf(*University Organization*)
ObjectPropertyDomain(*advisor Student*)
ObjectPropertyDomain(*headOf Dean*)
ObjectPropertyDomain(*takesCourse Student*)
FunctionalObjectProperty(*headOf*)
EquivalentClasses(*Person* ObjectUnionOf(*Student Professor*))
SubClassOf(*Student* ObjectSomeValuesFrom(*takesCourse Course*))
SubClassOf(*Professor* ObjectSomeValuesFrom(*worksFor University*))
SubClassOf(*Professor* ObjectSomeValuesFrom(*teacherOf Course*))
SubClassOf(*Dean* ObjectSomeValuesFrom(*headOf College*))
DisjointClasses(*Dean* ObjectSomeValuesFrom(*teacherOf Course*))

Fig. 15.3 The ontology of the example

faculty(*UNIVERSITY_CODE, CODE, DESCRIPTION*)
students(*ID, FNAME, SNAME, DOB, ADDRESS*)
course(*FACULTY_CODE, CODE, DESCRIPTION*)
assignment(*COURSE_CODE, PROFESSOR, YEAR*)
professor(*CODE, FNAME, SNAME, ADDRESS, PHONE*)
exam(*STUD_ID, COURSE_CODE, DATE, RATING*)
career(*STUD_ID, ACADEMIC_YEAR, FACULTY_CODE*)
degree(*STUD_ID, YEAR, PROF_ID, TITLE*)

Fig. 15.4 Relational tables of the source schema of the example

M is a set of mapping assertions, each one relating a query over the source schema to a query over the ontology.

Example. We now present an example of an OBDI system extracted from a real integration experiment involving data from different sources in use at Sapienza University of Rome. The ontology is defined by means of the OWL 2 assertions shown in Fig. 15.3.

It is, in fact, a portion of the Lehigh University Benchmark (LUBM) ontology, an ontology that is commonly used for testing ontology-based applications in the Semantic Web. In particular, the global schema contains the classes *Person*, *Student*, *Professor*, *Organization*, *College*, *Dean*, and *Course*, and the object properties *headOf*, *worksFor*, *takesCourse*, and *advisor*. For the sake of simplicity, we do not report in this example assertions involving data properties (i. e., attributes), but they are obviously allowed in our framework. The source schema is a set of relational tables resulting from the federation of several data sources of the School of Engineering of the Sapienza University of Rome, and the portion that we consider in this example is constituted by the relational tables shown in Fig. 15.4.

As for the mapping, referring to the global and source schemas presented above, we provide some sample mapping assertions in Fig. 15.5.

The mapping assertion M_1 specifies that the tuples from the source table *students* provide the information needed to build the instances of the class *Student*. In particular, the SQL query in the body of M_1 retrieves the code for students whose date of birth is before 1990; each such code is then used to build the object identifier for the student by means of the unary function symbol "st". Similarly, the mapping M_2 extracts data from the table *degree*, containing information on the student's Master's degree, such as the year of the degree, the title of the thesis, and the code of the advisor. The tuples retrieved by the query in the body of M_2, involving only degree titles earned after 2000, are used to build instances for the object property (relationship) advisor; the instances are constructed by means of the function symbols "pr" and "st". Finally, the mapping assertion M_3 contributes to the construction of the domain of the advisor, taking from the source table exam only codes of students who have passed the exam of courses that were not assigned to any professor before 1990.

```
M1: SELECT ID
    FROM students
    WHERE DOB <= '1990/01/01' }                      →  SELECT ?st(ID)
                                                        {?st(ID) rdf:type Student }
M2: SELECT STUD_ID, PROF_ID
    FROM degree  WHERE YEAR > 2000 }                 →  SELECT ?st(STUD_ID) ?pr(PROF_ID)
                                                        {?st(STUD_ID) advisor ?pr(PROF_ID)}
M3: SELECT STUD_ID
    FROM exam
    WHERE course CODE NOT IN
      (SELECT COURSE_CODE
       FROM assignment WHERE YEAR < 1990) }          →  SELECT ?st(STUD_ID)
                                                        { ?st(STUD_ID) advisor ?X }
```

Fig. 15.5 Mapping assertions of the example

An *OBDI system* is a pair (I, D), where I is an OBDI specification and D is a database for the source schema S, called source database for I. The semantics of (I, D) is given in terms of the logical interpretations that are models of O (i. e., satisfy all axioms of O, and satisfy M with respect to D). The notion of mapping satisfaction depends on the semantic interpretation adopted on mapping assertions. Commonly, such assertions are assumed to be sound, which intuitively means that the results returned by the source queries occurring in the mapping are a subset of the data that instantiate the ontology. The set of models of I with respect to D is denoted with ModD(I).

As we said before, in OBDI systems, the main service of interest is query answering, i. e., computing the answers to user queries, which are queries posed over the ontology. Such service amounts to return the so-called certain answers, i. e., the tuples that satisfy the user query in all the interpretations in ModD(I). Query answering in OBDI is thus a form of reasoning under incomplete information and is much more challenging than classical query evaluation over a database instance.

It is well known that carrying out inference tasks, such as answering queries in OBDI systems, may be computationally expensive. From the computational perspective, query answering depends on:

1. The language used for the ontology
2. The language used for user queries, and
3. The language used to specify the queries in the mapping.

In the following, we consider a particular instantiation of the OBDI framework, in which we choose each such language in such a way that query answering is guaranteed to be tractable with respect to the size of the data.

From the general framework we obtain a computationally tractable one by choosing appropriate languages as follows:

- The ontology language is *DL-Lite*$_{\mathrm{A}}$ or its subset *DL-Lite*$_{\mathrm{R}}$ [15.22].
- The mapping language follows the *global-as-view* (GAV) approach [15.6].
- The user queries are unions of conjunctive queries [15.16].

In the following, we discuss each of the above choices.

15.3.1 Ontology Language

DL-Lite$_{\mathrm{A}}$ [15.22] is essentially the maximally expressive member of the *DL-Lite* family of lightweight DLs [15.20]. In particular, its subset *DL-Lite*$_{\mathrm{R}}$ has been adopted as the basis of the OWL 2 QL profile of the W3C standard OWL. As usual in DLs, *DL-Lite*$_{\mathrm{A}}$ allows for representing the domain of interest in terms of *concepts*, denoting sets of objects, and *roles*, denoting binary relations between objects. In fact, *DL-Lite*$_{\mathrm{A}}$ also considers *attributes*, which denote binary relations between objects and values (such as strings or integers), but for simplicity, we do not consider them in this chapter. From the expressiveness point of view, *DL-Lite*$_{\mathrm{A}}$ is able to capture essentially all the features of entity-relationship diagrams and UML (Unified Modeling Language) class diagrams, except for completeness of hierarchies. In particular, it allows for specifying ISA ("is a") and disjointness between either concepts or roles, mandatory participations of concepts into roles, and the typing of roles. Formally, a *DL-Lite*$_{\mathrm{A}}$ TBox is a set of assertions obeying the syntax

$B_1 \sqsubseteq B_2$	(positive concept inclusion)
$B_1 \sqsubseteq \neg B_2$	(negative concept inclusion)
$R_1 \sqsubseteq R_2$	(positive role inclusions)
$R_1 \sqsubseteq \neg R_2$	(negative role inclusions)
(funct R)	(role functionalities)

where:

- B_1 and B_2 are basic concepts, i. e., expressions of the form A, $\exists P$, or $\exists P^-$.
- R, R_1, and R_2 are a basic roles, i. e., expressions of the form P, or P^-.
- A and P denote an *atomic concept* and an *atomic role*, respectively, i. e., a unary and binary predicate from the ontology alphabet, respectively.
- P^- is the *inverse* of an atomic role P, i. e., the role obtained by switching the first and second components of P.
- $\exists P$ (or $\exists P^-$), called existential unqualified restriction, denotes the projection of the role P on its first (or second) component.
- $\neg B_2$ (or $\neg R_2$) denotes the negation of a basic concept (or role).

Assertions of the form (funct R) are called role functionalities and specify that an atomic role, or its inverse, is functional. *DL-Lite*$_A$ poses some limitations on the way in which positive role inclusions and role functionalities interact. More precisely, in a *DL-Lite*$_A$ TBox an atomic role that is either functional or inverse functional cannot be specialized, i. e., if (funct P) or (funct P^-) are in the TBox, no inclusion of the form $R \sqsubseteq P$ or $R \sqsubseteq P^-$ can occur in the TBox. *DL-Lite*$_R$ is the subset of *DL-Lite*$_A$ obtained by removing role functionalities altogether.

A *DL-Lite*$_A$ *interpretation* $J = (\Delta^J, \cdot^J)$ consists of a non-empty set constituting the *interpretation domain* Δ^J and an *interpretation function* $\cdot^J$ that assigns to each atomic concept A a subset A^J of Δ^J, and to each atomic role P a binary relation P^J over Δ^J. In particular, for the constructs of *DL-Lite*$_A$, we have (symbol \ is used to denote set difference):

- $A^J \subseteq \Delta^J$
- $P^J \subseteq \Delta^J \times \Delta^J$
- $(P^-)^J = \{(o_2, o_1) | (o_1, o_2) \in P^J\}$
- $(\exists R)^J = \{o | \exists o'.(o, o') \in R^J\}$
- $(\neg B)^J = \Delta^J \setminus B^J$
- $(\neg R)^J = (\Delta^J \times \Delta^J) \setminus R^J$

Let C be either a basic concept B or its negation $\neg B$. An interpretation J satisfies a concept inclusion $B \sqsubseteq C$ if $B^J \subseteq C^J$, and similarly for role inclusions. Also, J satisfies a role functionality (funct R) if the binary relation R^J is a function, i. e., $(o, o_1) \in R^J$ and $(o, o_2) \in R^J$ implies $o_1 = o_2$.

15.3.2 Mapping Language

The mapping language in the tractable framework allows mapping assertions of the following the forms:

- $\varphi(x) \rightarrow A(f(x))$
- $\varphi(x) \rightarrow P(f_1(x_1), f_2(x_2))$

where $\varphi(x)$ is a domain-independent first-order query (i. e., an SQL query) over S, with free variables x, A and P are as before variables in x_1 and x_2 also occur in x, and f, possibly with subscripts, is a function.

Intuitively, the mapping assertion of the first form, called the concept mapping assertion, specifies that individuals that are instances of the atomic concept A are constructed through the use of the function f from the tuples retrieved by the query $\varphi(x)$. Similarly for the mapping assertion of the second form called the role mapping assertion. Each assertion is of type GAV [15.6], i. e., it associates a view over the source (represented by $\varphi(x)$) to an element of ontology. However, differently from traditional GAV mappings, the use of functions is crucial here, since we are considering the typical scenario in which data sources do not store the identifiers of the individuals that instantiate the ontology, but only maintain values. Thus, functions are used to address the semantic mismatch existing between the extensional level of S and O [15.22].

Formally, we say that an interpretation J satisfies a mapping assertion $\varphi(x) \rightarrow A(f(x))$ with respect to a source database D, if for each tuple of constants t in the evaluation of $\varphi(x)$ on D, $(f(t))^J \in A^J$, where $(f(t))^J \in \Delta^J$ is the interpretation of f(t) in J that is, f(t) acts simply as a constant denoting an object. Satisfaction of assertions of the form $\varphi P(f_1(x_1), f_2(x_2))$ is defined analogously. We also point out that *DL-Lite*$_A$ adopts the unique name assumption (UNA), that is, different constants denote different objects, and thus different ground terms of the form f(t) are interpreted with different elements in Δ^J.

15.3.3 User Queries

In our tractable framework for OBDI, user queries are conjunctive queries (CQs) [15.16] or unions thereof. With q(x), we denote a CQ with free variables x. A Boolean CQ is a CQ without free variables. Given an OBDI system (I,D) and a Boolean CQ q over I, i. e., over the TBox of I, we say that q is *entailed by* (I, D), denoted with $(I, D) \models q$, if q evaluates to true in every interpretation $J \in \mathrm{Mod}_D(I)$. When the user query $q(x)$ is non-Boolean, we denote with $\mathrm{cert}_D(q(x), I)$ the *certain answers* to q with respect to (I, D), i. e., the set of tuples t such that $(I, D) \models q(t)$, where q(t) is the Boolean CQ obtained from q(x) by substituting x with t.

15.3.4 Query Answering

Although query answering in the general framework may soon become intractable or even undecidable, depending on the expressive power of the various languages involved, the tractable framework has been designed to ensure tractability of query answering. We end this section by illustrating the basic idea to achieve tractability.

In the tractable OBDI framework previously described, one can think of a simple chase procedure [15.28] for query answering, which first retrieves an initial set of concept and role instances from the data source through the GAV mapping, and then, using the ontology axioms, *expands* such a set of instances deriving and materializing all the logically entailed concept and role assertions; finally, queries can be evaluated on such an expanded set of instances. Unfortunately, in $\textit{DL-Lite}_A$ (and in $\textit{DL-Lite}_R$ already) the instance materialization step of the above technique is not feasible in general, because the set of entailed instance assertions starting from even very simple OBDA specifications and small data sources may be infinite.

As an alternative to the above materialization strategy, most of the approaches to query answering in OBDI are based on query rewriting, where the aim is to first compute a query q' representing a reformulation of a query q with respect to an OBDI specification I, and then evaluate q' over the source database. Actually, the above described OBDI framework allows for modularizing query rewriting. Indeed, the current techniques for OBDI consist of two phases, namely the phase of *query rewriting with respect to the ontology*, and the phase of *query rewriting with respect to the mapping*:

1. In the first phase, the initial query q is rewritten with respect to the ontology, producing a new query q_1, still over the ontology signature; intuitively, q_1 *encodes* the knowledge expressed by the ontology that is relevant for answering the query q.
2. In the second phase, the query q_1 is rewritten with respect to the mapping M, thus obtaining a query q_2 to be evaluated over the source data. Thus, the mapping assertions are used for reformulating the query into a new one expressed over the source schema signature.

Thus, following the above method, computing the certain answers of a query Q over I is reduced to a simple evaluation of a suitable query over the source database, thus relying on the technology of relational databases, including the possibility of adopting well-established query optimization strategies.

15.4 *Sapientia* and OBDI for Multidimensional Research Assessment

Sapientia is the ontology of multidimensional research assessment developed by an interdisciplinary group of scholars funded by Sapienza University of Rome in the framework of its Research Awards (two main research projects). It models all the activities relevant for the evaluation of research and for assessing its impacts. For impact, in a broad sense, we mean any effect, change or benefit, to the economy, society, culture, public policy or services, health, the environment, or quality of life, beyond academia.

The first version of *Sapientia* was closed on the 22 December 2014. It consisted of 14 modules of around 350 concepts, roles and attributes (Fig. 15.6). Figure 15.7 illustrates the organization of the main components of an OBDI system including *Sapientia*.

15.4.1 *Sapientia*'s Philosophy and its Main Principles

Sapientia's mission is being comprehensive and try to model everything related to the evaluation of research and its impacts. As we stated in Sect. 15.1, the evaluation of research, is a complex activity. It requires a systematic view and has its base on the interplay between theory, methodology, and data. To accomplish its mission, we designed *Sapientia* to be at the heart of a flexible knowledge infrastructure ("robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds" [15.29]) for the multidimensional assessment of research and its impacts. By flexible we mean a knowledge infrastructure characterized by being an open and evolving infrastructure.

The principles followed in the design and development of *Sapientia* are the following:

- We started with a top-down modeling approach, with subsequent bottom-up refinements and cyclical improvements. We describe and model the domain from a conceptual point of view, without considering the existing data and its specificity.

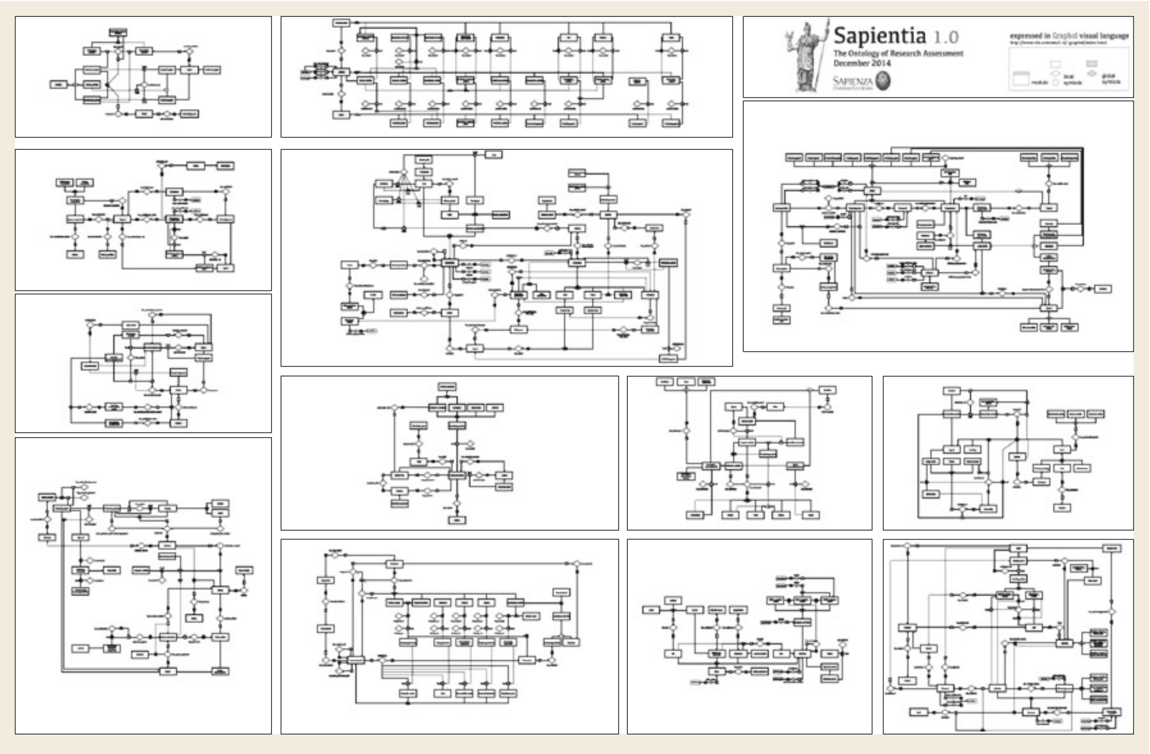


Fig. 15.6 An illustration of *Sapientia* (the ontology of multidimensional research assessment) 1.0

- We left outside the scope of the ontological commitment all the methodological consideration about choice of the methods for the assessment of research. This is because we want that our ontology being the common ground for experimenting and testing different methods and approaches.
- We left outside the scope of the ontological commitment the implementation problem and the consequences of evaluation. Again, this is for keeping our ontology as a common ground, a shared language or vocabulary, to build a cooperative and open discussion about evaluation approaches considering the interaction of different stakeholders with different points of view and interests.
- We pursued a modeling approach based on processes, which are conceived as collections of activities. A process is composed by inputs and outputs.
- Individuals and activities are the main pillars of the ontology.
- We followed a modeling approach based on a modularization of the system. Our ontology is organized in modules. As we shall see later, we have two kind of modules: functional modules and structural modules. By functional modules we mean modules that model the main agents and activities of our domain (namely Agents, Activities, R&D, Publishing, Education, Resources and Review). By structural modules, we mean those modules that represent the constituent elements of the ontology to ensure its long lasting and general-purpose functionality (namely, Taxonomies, Space, Representations, and Time).

We consider the building of descriptive, interpretative, and policy models of our domain as a distinct step with respect to the building of the domain ontology. In the following part of this section, we will reproduce the findings that we previously published in *Daraio*, *Lenzerini*, et al. [15.24]. However, the ontology will intermediate the use of data in the modeling step and should be rich enough to allow the analyst the freedom to define any model she considers useful to pursue her analytic goal.

Obviously, the actual availability of relevant data will constrain both the mapping of data sources on the ontology and the actual computation of model variables and indicators of the conceptual model. However, the analyst should not refrain from proposing the models that she considers the best suited for her purposes and to express, using the ontology, the quality requirements, the logical, and the functional specification for her ideal model variables and indicators. This approach has many merits, in particular:

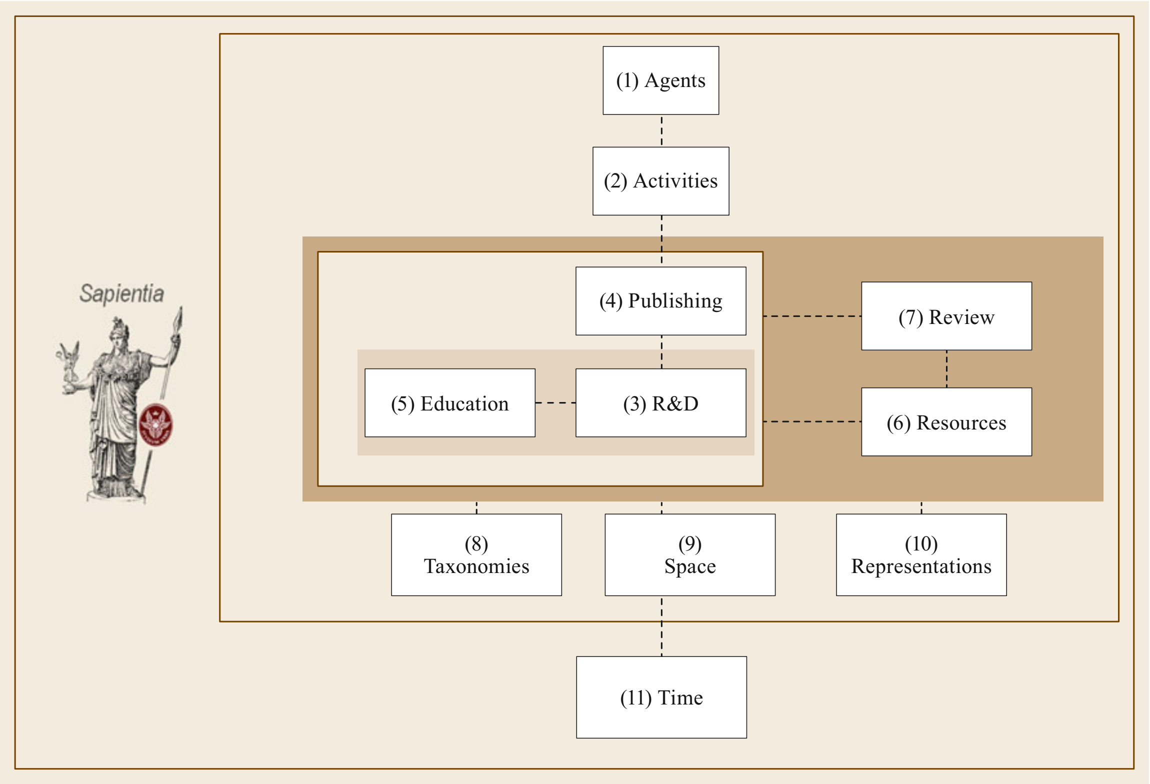


Fig. 15.7 Modules of Sapientia 3.0

- It allows the use of a common and stable ontology as a platform for different models.
- It addresses the efforts to enrich data sources, and verify their quality.
- It makes transparent and traceable the process of approximation of variables and models when the available data are less than ideal; it makes use of every source at the best level of aggregation, usually the atomic one.

More generally, this approach is consistent with the effort of avoiding "the harm caused by the blind symbolism that generally characterizes a hasty mathematization" put forward by *Georgescu-Roegen* in his seminal work on production models and on methods in economic science [15.30–32]. In fact, one can verify the logical consistency of the ontology and compute answers to unambiguous logical queries.

Moreover, the proposed ontology allows us to follow the Georgescu-Roegen approach also in the use of the concept of process. We can analyze the knowledge production activities at an atomic level, considering their time dimension and such funds as the cumulated results of previous research activities, both those available in relevant publications and those embodied in the authors' competences and potential, the infrastructure assets, and the time devoted by the group of authors to current research projects. Similarly, we can analyze the output of teaching activities, considering the joint effect of funds such as the competence of teachers, the skills and initial education of students, and educational infrastructures and resources. Thirdly, service activities of research and teaching institutions provide infrastructural and knowledge assets that act as a fund in the assessment of the impact of those institutions on the innovation of the economic system. The perimeter of our domain should allow us to consider the different channels of transmission of that impact: mobility of researchers, career of alumni, applied research contracts, joint use of infrastructures, and so on. In this context, different theories and models of the system of knowledge production could be developed and tested.

15.4.2 Requested Investment and Modularity of the System

In an OBDI approach, the modular design and its implementation requires an initial *large scale investment* into the formal definition of the main relevant concepts (and relationships among them) of the domain of interest but is facilitated by suitable graphical tools (which we will see below) that allow an easy modularization and updates of the relevant domain. The following paragraph is taken from *Daraio*, *Lenzerini*, et al. [15.33].

Following a real options approach in investment theory [15.34], we can conceive a data platform as an asset allowing repeated use. In this context, investment costs are made by front-up costs for the platform, maintenance costs, and recurring costs for projects. The revenues instead are the gains from better decisions in policy making (e.g., the possible use for performance-based allocation of public resources, the possible use for strategic priorities in S&T, or to set up public subsidies to firms for industrial R&D). A real options analysis in this context should follow a modular engineering design perspective [15.35] in which a quantitative model to describe the economic forces that push a design towards modularization and the consequences of modularity on the business environment are described. In this context, value creation is the goal of the modularization process, and real options theory offers a natural framework to evaluate the modularization of the design of the system.

There are also criteria to assess the decomposition of systems into modules. In modular design, the main criteria to assess the decomposition of systems into modules are those of cohesion and coupling. The principal rule to assess the quality of the modularization of a system, attributed to *Parnas* [15.36] even if the paper does not contain the terms cohesion and coupling, is of *high cohesion within modules and loose coupling between modules*.

Cohesion refers to the degree to which the elements of a module belong together and, hence, it is a measure of how strongly related each piece of a module is. Modules with high cohesion tend to be preferable, because high cohesion is associated with several desirable properties including robustness, reliability, reusability, and understandability.

Coupling is the manner and degree of interdependence between modules; a measure of how closely connected two modules are. Low coupling is often a sign of a well-structured system and a good design, and when combined with high cohesion, supports the general goals of high readability and maintainability. Modularity is a property of quasi-decomposition of hierarchical systems, based on the minimization of the interdependence of subsystems [15.37].

The modification of subsystems does not require the re-design of the entire system. Making the design of products modular requires a large front up investment in conceptual design. The standardization of interfaces is necessary. However, the design of successive versions of the product and/or re-design becomes cheaper.

The current version of *Sapientia*, *Sapientia* 3.0, includes around 600 concepts, roles, and attributes and is organized in 11 modules (Fig. 15.7). Another module on skills is going to be included. Its aim is to model all the competences involved in the assessment of research and its impacts.

The *central* modules of *Sapientia* are Agents (no.1) and their Activities (no.2), which are expanded into the five *main* process-based modules: 3 R&D, 4 Publishing (ancillary module of R&D), 5 Education, 6 Resources, 7 Review. These are connected to the four *auxiliary* modules (also defined as *structural* modules): 8 Taxonomies, 9 Space, 10 Time and 11 Representations, which support the modeling of the main modules from 1 to 7.

The module Agents (1) models the subjects involved in the world of research, carrying out the activities described in module 2 (Activities). Activities models, overall, the relevant actions carried out by agents and their products. Module 3 (R&D) models Research and Development (R&D) activities, those that allow the scientific community to advance the state of the art of knowledge. The module Publishing (4) models the publishing activities, those that allow people to communicate (and disseminate) the results of the R&D activities carried out. Education (5) models the educational activities, those that allow people to improve their knowledge and to acknowledge the improvements made by other people. Resources (6) models the resources and the activities carried out for their management. Review (7) models the reviewing activities for assessing the R&D activities and their results. Taxonomies (8) models the nomenclatures that classify the several elements of the domain. Space (9) models the regions of space where agents and activities are located and their roles. Representations (10) models the representations of the objects of the domain according to a Source. A data source is a possible source, but also other sources, such as a theory may be a source. Time (11) models the depth of time of the domain and cut across all the other modules.

15.5 Reasoning over *Sapientia*: Some Illustrative Examples

Sapientia, the ontology introduced in the previous section, is an important element of an OBDI system introduced in the previous sections (Fig. 15.8).

An OBDI system allows us to reason on each component of the system, i. e., on the ontology, on the data sources, and on the mappings. In the following, we show some examples.

15.5.1 Reasoning over the Ontology

In this section, we illustrate an example of reasoning over the ontology *Sapientia* analyzing some extracts from the Module 3 (R&D). The R&D module aims at modeling the research activities of researchers and their products. The central concept of the module is the R&D activity that is linked to two important concepts, i. e., research product and research outcome.

In *Sapientia*, any research activity has:

- Its direct output (has_output), available without the contributions of any other activity.
- Its outcome (has_outcome): an output of any activity (not necessarily a research activity) participating a value chain where the research activity has an enabling role (i. e., without the research activity, that specific output would not be generated).

In *Sapientia*, a Publication aims at reporting empirical or theoretical work and describes the results obtained in some knowledge field. Publications are described in the Module Publishing (4), which concerns the activity that allows people knowing the results of research. The output of a publishing activity is a publication, which is a way to represent a content through some media. There are four kinds of contents in *Sapientia*: paper-like content (a content structured as for being published paper); book-like content (a content structured as for being published as monographs or edited chapters), patent-like content (a content structured as for being published as patent applications) and Project-like content (a content structured as being suitable to apply for a call).

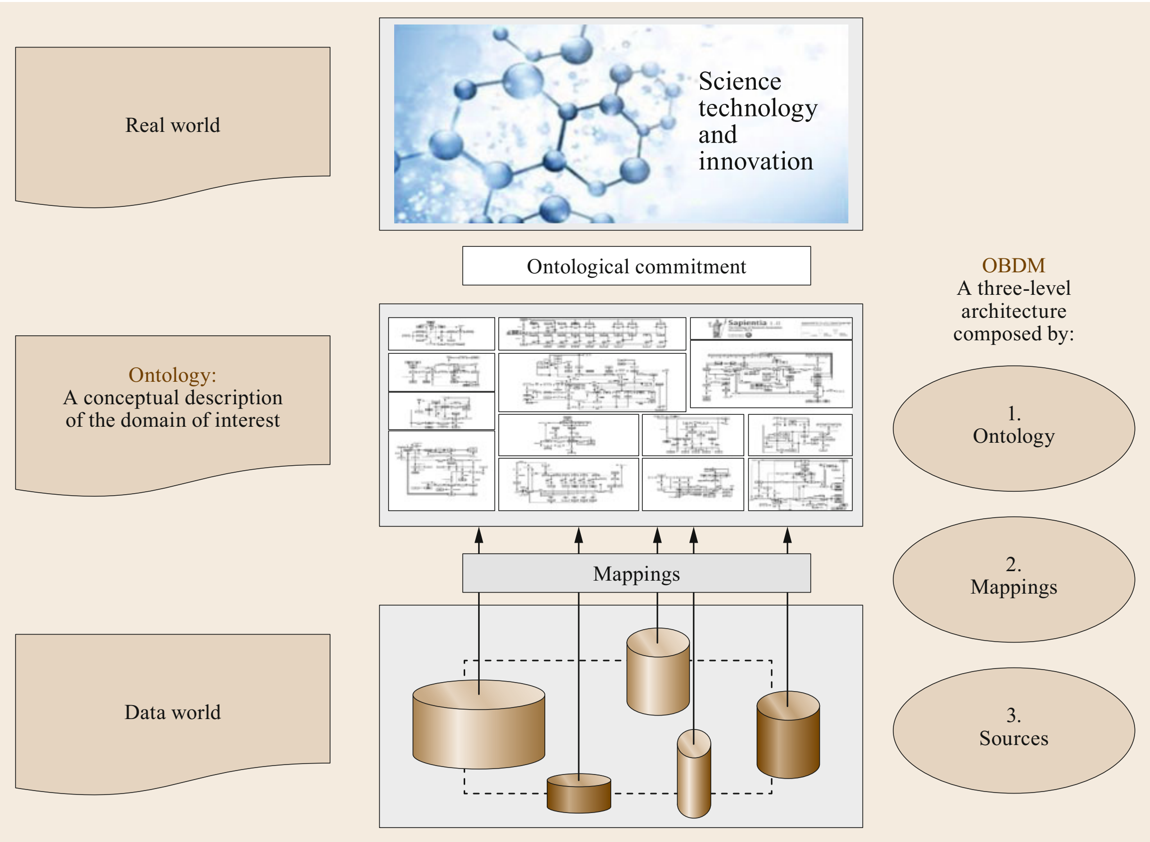


Fig. 15.8 Key components of an OBDI information system (after [15.15])

We point out that publication of research work is not considered an output of that research, because the output of the research work is the content of the publication. Publications, in fact, are the main concept of another module, the Module 4 (Publishing).

There are three kinds of agents involved in a publication:

- *Author*: An author of a publication is an agent that has contributed in writing the content of the publication (for instance, reporting the results of some research (s)he has carried out).
- *Editor*: An editor of a complex publication (where contributions of different authors need to be verified, harmonized, and combined) is an agent that oversees and coordinates the publication.
- *Publisher*: A publisher of a publication is the agent that provides some media to deliver and display a publication.

In *Sapientia* there are three kinds of publications:

- *Atomic publications*: A publication resulting from a unique, indivisible act of writing by one or more authors.
- *Collections*: A publication disseminating a group of atomic publications in a unique impulse, during a limited and short period of time.
- *Series*: Each disseminating a group of atomic publications during a long and (perhaps) unlimited period of time.

In *Sapientia*, a patent application is a possible publication, and is the output of an applied research. Notice that a patent is different from the other types of atomic publication; it is a right granted by a state which may concern a research output, not an output itself. A patent application follows its own path within the three levels of publications:

- It is an atomic publication itself.
- It is published in an issue (a collection).
- That issue appears in an intellectual property law journal (a series).

Note that there are no constraints between contents and publications where they can be published (for example, a patent_like_content can be placed in a part of a paper).

In Figs. 15.9 and 15.10, we show an example extracted from the Module R&D of *Sapientia* 3.0. Figures 15.9 and 15.10 display the path from Researcher to Publication in *Sapientia*. Figure 15.11 reproduces a legend to interpret the symbols used in the previous Figs. 15.9 and 15.10. Table 15.1 describes the main concepts and relations showed in Figs. 15.9 and 15.10. The language Graphol (http://www.dis.uniroma1.it/~graphol/) developed at the Sapienza university and implemented in the software Eddy [15.38, 39] is used in Figs. 15.9 and 15.10.

Graphol permits the expression of the axioms of an ontology as described in Sections 2 and 3 but in a more readable manner, in the form of a graph. This graphical representation of an ontology is very practical and useful to those who do not have a thorough knowledge of mathematical logic. Using the editor Eddy, it is possible to construct the corresponding chart ontology expressed in Graphol, which can be automatically translated (with a suitable translator downloaded from the website of Graphol) in a superset of OWL, or in a set of axioms OWL, possibly with the addition of some axioms that are not directly expressible in OWL (such as those of identification and denial of *DL-lite*). The graph expressed by Graphol illustrates and highlights the relationship between the various concepts and the various reports. The purpose of the graph is to offer a schematic view of the ontology, to focus attention on the concepts and how they are mutually linked in the representation.

The examination of Figs. 15.9 and 15.10 allows us to reason over the ontology about the path from researcher to publication. It clearly appears that the path from researcher to R&D activities to publication goes through *content*.

15.5.2 Reasoning over the Mappings

The mappings in an OBDI system play a crucial role. They bind the data sources to the ontology. The connection is made through the *materialization or staging* phase (Fig. 15.12). Different levels of materialization are possible, and the refreshing of the materialization can be different according to the data source. The realization of the mappings in the context of *Sapientia* and its OBDI system is an interesting peculiar case, because the problem of *disambiguation* is a very important problem in bibliometrics and affects the assessment of the research and its impacts.

The solution proposed so far for the implementation of the mappings of *Sapientia* to its data sources is based on the *balance* between the level of computation and the level of materialization. Therefore, *quality checks* on data sources can be carried out.

The connection with the entity resolution (ER) occurs at the *computation* level through *modularization*. This allows a high degree of flexibility; ER algorithms can be replaced and/or updated without modifying the overall system. This is a *computational* flexibility

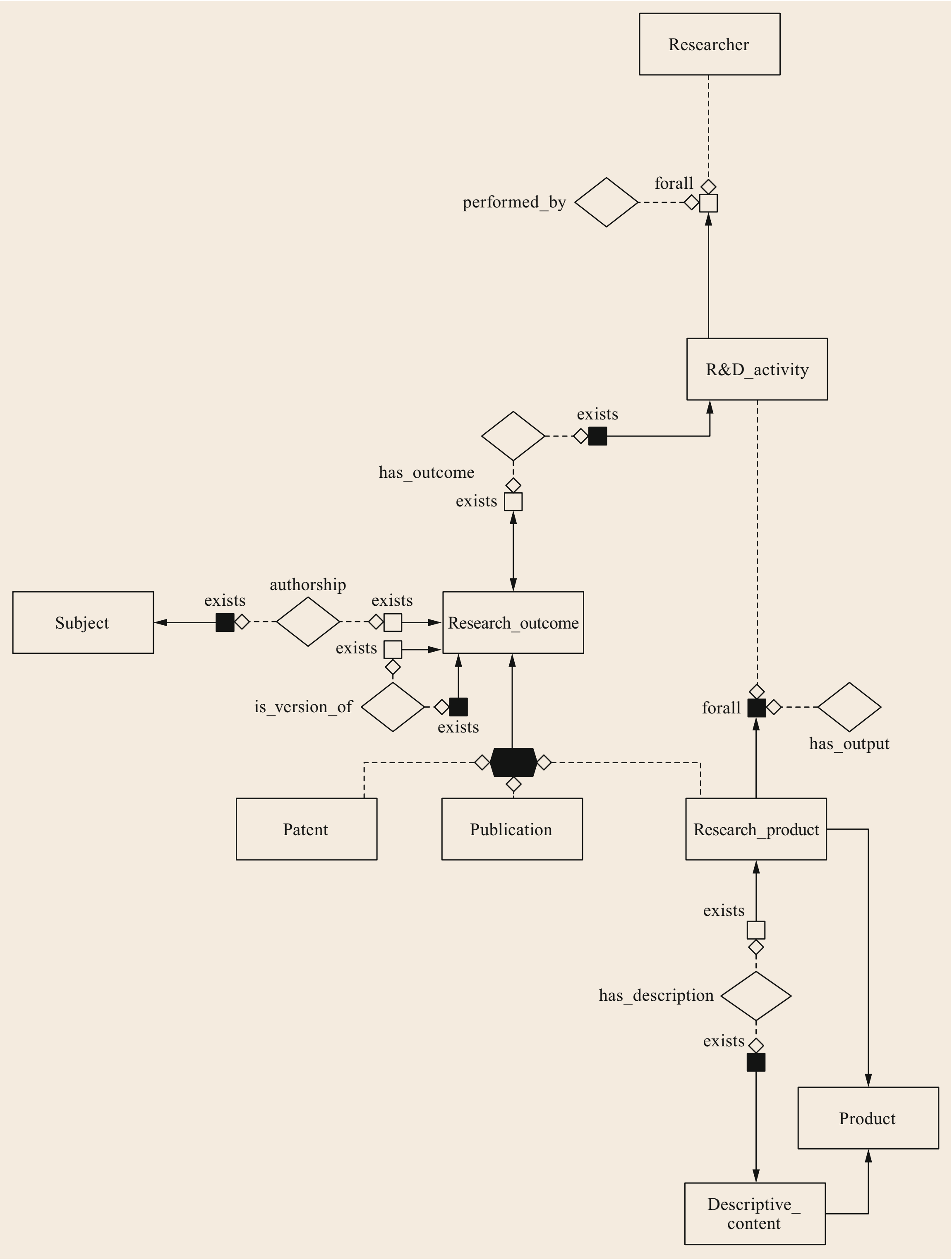


Fig. 15.9 An illustration of *Sapientia* from Module 3 R&D. Part I: from researcher to descriptive content

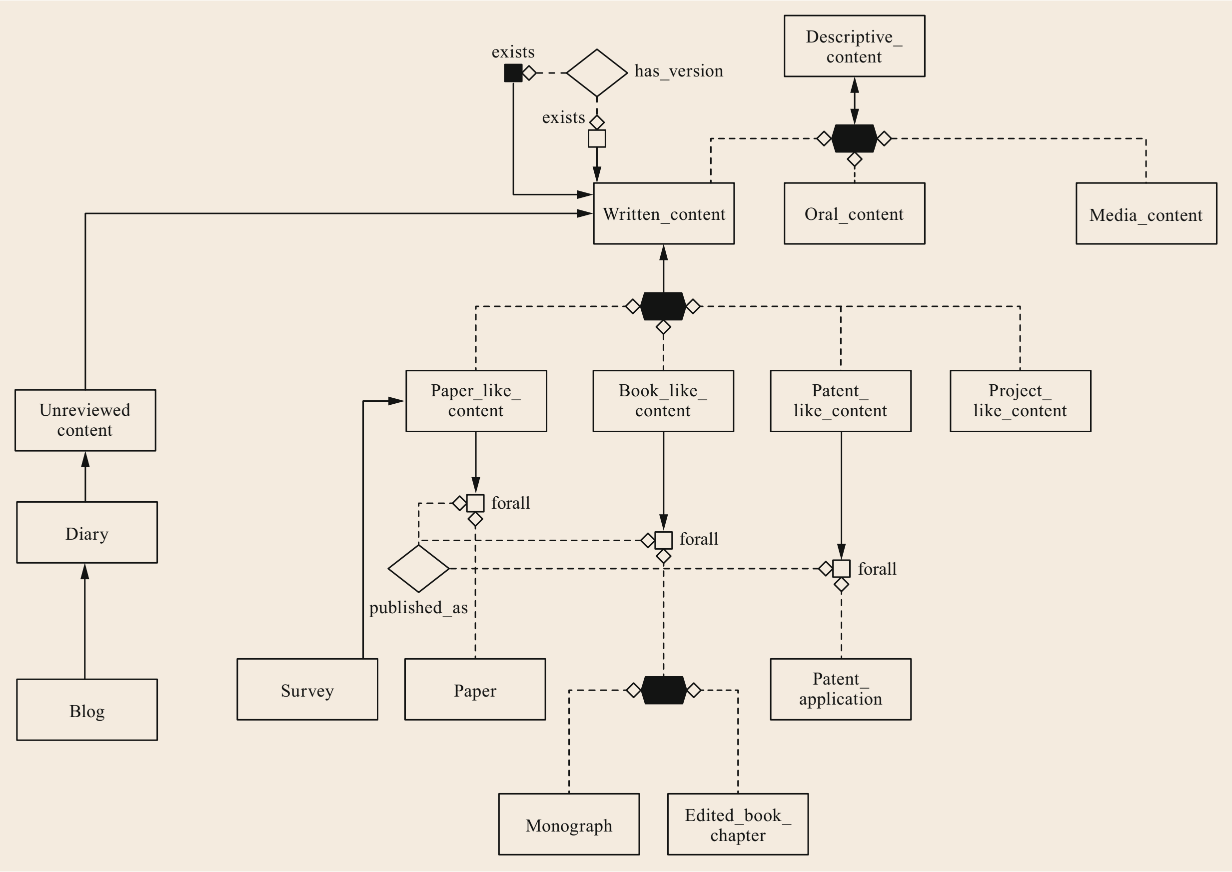


Fig. 15.10 An illustration of *Sapientia* from Module 3 R&D. Part II: from descriptive content to publications

with ER techniques through *modularization*. In addition, the ontological model offers the advantage of analyzing authors and affiliations together by means of a *conceptual* flexibility through the Module 11 Representations.

So far, we have implemented blocking mechanisms, mainly based on the Inverted Index. As with pairs of comparisons, blocking mechanisms are based on a common interface and specific implementations for single data sources. They act in a *spiral model* that will allow us to refine ER algorithms in subsequent iterations. In this respect, the ontology can be supportive to identify the strengths and weaknesses of the different ER algorithms that will be tested and implemented to our domain.

De Giacomo et al. [15.14] in the current challenges for OBDI mention “the problem of devising methodologies and tools for developing mappings for OBDI” and state that it “is largely unexplored.”

15.5.3 Reasoning over the Data and Indicators

Other possible reasoning tasks are those on data sources and on the indicators that may be built over the data. *Daraio*, *Lenzerini*, et al. [15.33] showed the usefulness of an OBDI system for R&I integration from a data quality perspective. *Daraio*, *Lenzerini,* et al. [15.42] showed that an OBDI approach allows for an unambiguous specification of indicators according to its four main dimensions: ontological, logical, functional, and qualitative (Table 15.2).

Reasoning over *Sapientia* permits the characterization of each indicator along its four dimensions listed above. This is extremely useful for studying the specification of indicators in the context of an *ontological approach* to the evaluation of research. The specification of the indicators within an OBDI framework aims to protect the analyst from the *risk of reductive conclu-*

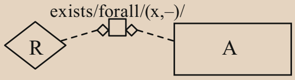
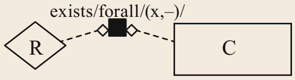
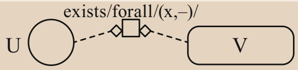
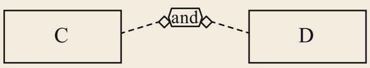
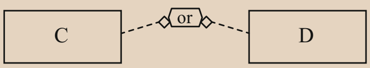
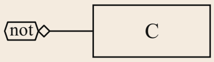
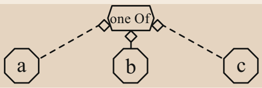
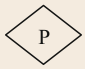
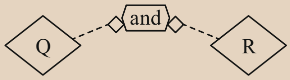
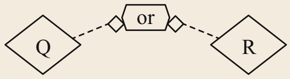
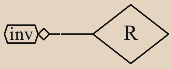
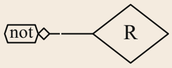
Expression	DL-syntax	Graphol syntax
Atomic concept	A	
Domain restriction on role	$\exists R.C \quad \forall R.C$ $\geq xR.C \quad \leq yR.C$	
Range restriction on role	$\exists R^{-}.C \quad \forall R^{-}.C$ $\geq xR^{-}.C \quad \leq yR^{-}.C$	
Domain restriction on attribute	$\exists U.V \quad \forall U.V$ $\geq xU.V \quad \leq yU.V$	
Concept intersection	$C \sqcap D$	
Concept union	$C \sqcup D$	
Concept complement	$\neg C$	
Concept one-of	$\{a, b, c\}$	
Atomic role	P	
Role intersection	$Q \sqcap R$	
Role union	$Q \sqcup R$	
Role inverse	R^{-}	
Role complement	$\neg R$	
Role chain	$Q \circ R$	

Fig. 15.11 Legend of the symbols illustrated in Figs. 15.9 and 15.10 (after [15.38, 39])

sions, which would focus on the logical and functional aspects of the specification, ignoring the ontological and quality assessment parts of the process.

In a broader perspective, *Daraio* and *Bonaccorsi* [15.43] identify two trends in indicator development:

- Trend towards *granularity* of indicators ("new indicators are explicitly requested to allow various kinds of aggregation and disaggregation, preserving desirable statistical properties, in order to address new policy needs"), detailing granularity in territorial, institutional and disciplinary areas.
- Trend towards *cross-referencing* ("the ability of indicators to be combined in meaningful ways, preserving their statistical properties").

Sapientia and its OBDI system may be a suitable infrastructure to develop indicators that satisfy the policy requirements of granularity and cross-referencing in a coherent and consistent way.

Table 15.1 Main concepts and roles illustrated in Figs. 15.9 and 15.10

Term	Type	Definition
Researcher	Concept	A researcher is an agent that carries out research activities allowing the scientific community to advance the state of the art of knowledge
R&D activity	Concept	A research and experimental development (R&D) activity, according to the Frascati manual [15.40, p. 28] "comprise creative and systematic work undertaken in order to increase the stock of knowledge—including knowledge of humankind, culture and society—and to devise new applications of available knowledge."
Research_product	Concept	A research product is a product that is an output of a research activity. In *Sapientia*, we included all the research products of the *Research Excellence Framework* [15.41].
Research_outcome	Concept	A research outcome of a research activity R is the output of an activity (not necessarily a research activity) participating a value chain where R has an enabling role (without R that output would not be generated).
Subject	Concept	A subject is any entity that can act as an agent and, playing such role, performs some activities. There are two types of subjects: natural persons and organizations.
Patent	Concept	A patent is a right that may be owned in an ownership, and shall confer to its owner the following exclusive rights: where the subject matter of a patent is a product, to prevent third parties not having the owner's consent from the acts of: making, using, offering for sale, selling, or importing for these purposes that product where the subject matter of a patent is a process, to prevent third parties not having the owner's consent from the act of using the process, and from the acts of using, offering for sale, selling, or importing for these purposes at least the product obtained directly by that process. Patent owners shall also have the right to assign or transfer by succession the patent and to conclude licensing contracts (Article 28 of the Trade-Related Intellectual Property Rights (TRIPS) Agreement administered by the World Trade Organization (WTO) that sets down minimum standards for many forms of intellectual property). Patents grant their owner a set of rights of exclusivity over an invention (a product or process that is new, involves an inventive step and is susceptible of industrial application). The legal protection conferred by a patent gives its owner the right to exclude others from making, using, selling, offering for sale or importing the patented invention for the term of the patent, which is usually 20 years from the filing date, and in the country or countries concerned by the protection. This set of rights provides the owner with a competitive advantage. Patents can also be licensed or used to help create or finance a spin-off company. It is therefore possible to derive value from them even if their owner does not have its own manufacturing capability (e. g., universities).
Publication	Concept	A publication is a particular kind of product consisting of an atomic or complex media including some content. There are three kinds of publications: atomic, collections and series.
Product	Concept	A product is an output of an activity, an entity (that might satisfy a want or need expressed by someone—or something—different from the agent who carried out the activity) which appears in the domain as consequence of an activity. Notice that the activity does not need to be finished at the time one of its products appears.
Descriptive content	Concept	A descriptive content is an interpretable object from which a human or an artificial intelligence can capture a meaning. It can use linguistic expressions and or media content.
Written content	Concept	A written content is a set of resources suited to be included in a single publication. These can be texts, technical drawings, diagrams, photographs and so on. These resources come together with their organizations (chapters, paragraphs, index). The linguistic parts of the resources are written in one or more natural languages (has_language).
Oral content	Concept	A descriptive content is oral if is the content of a speech.
Media content	Concept	A descriptive content is a media content whether it represents the way one or more events stimulate human sight and/or hearing.
Paper-like content	Concept	A paper like content is a written content suitable to become a paper, with respect to its internal organization, its length, its illustrations (technical drawings and diagrams, for example) and so on. Every paper has its content but not every paper-like content is published as a paper (for example it exists before the publication).
Book-like content	Concept	A book like content is a written content suitable to be published as (a part of) a book, with respect to its internal organization, its length, its illustrations (technical drawings and diagrams, for example) and so on. Every book has its content but not every book-like content is published as a book (for example it exists before the publication).

Table 15.1 (continued)

Term	Type	Definition
Patent-like content	Concept	A patent like content is the a written content suitable to a patent application, with respect to its internal organization, its length, its illustrations (technical drawings and diagrams, for example), and so on. Every application has its content but not every patent-like content is published as an application (for example, if it exists before the application).
Project-like content	Concept	A project-like content is written content suitable to apply to answer a call with respect to its internal organization, its length, its illustrations (technical drawings and diagrams, for example), and so on. A project like content is not, in general, an object of publication.
Survey	Concept	A paper-like content is a survey if it is an attempt to summarize the current state of understanding on a topic or a knowledge area.
Paper	Concept	A paper is an atomic publication. It contains original research results or reviews existing results. Before publication, the content of the paper has undergone a process of *peer review* by one or more referees (who are experts of the same field) who have checked that the content of the paper is suitable for publication in the journal. Such content may undergo a series of reviews, revisions, and re-submissions before finally being accepted or rejected for publication.
Monograph	Concept	A monograph is an atomic publication. Although a monograph has, in general, a structure, all its parts share the same group of authors. A monograph can be physically distributed in more volumes.
Edited_book_chapter	Concept	An edited book chapter is an atomic publication that is a part of an edited book with specific authors.
Unreviewed_content	Concept	It is a written content that has not been reviewed.
Diary	Concept	A diary is a record (possibly in handwritten format) with discrete entries arranged by date reporting (typically) on what has happened over the course of a day or a period.
Blog	Concept	A blog (a truncation of the expression weblog) is a discussion or informational site published on the World Wide Web and consisting of discrete entries (*posts*) typically displayed in reverse chronological order (the most recent post appears first). Blogs often cover a single subject. A blog is a diary, since the posts are arranged by date.
Published_as	Relation	Binds a written content to the publication which disseminates it (if any)
Has_version	Relation	Binds a written contents to its new versions (if any)
Has_description	Relation	Binds a research activity with descriptive content that descriptive it (if the activity has descriptions); notice that the descriptive content of a research activity may not be one of its outputs, and an output of a research activity may not be a description of the activity itself
Has_output	Relation	Binds any activity to its outputs; an activity may produce its outputs at any time when it is operative.
Authorship	Relation	Binds a research outcome to the subject which is responsible for it
Has_outcome	Relation	Binds a research activity to any output of any activity (not necessarily a research activity) participating in a value chain where the research activity has an enabling role (without the research activity that output would not be generated). The figure shows chains schemes that justify the outcome of the research activity shown in the left. The arrows represent the role has_output.
Performed_by	Relation	Binds an activity to the agent who performs it

Table 15.2 Main indicator dimensions and their role in the process of indicator development

Dimension	Description	Role
Ontological	Conceptual characterization (knowledge representation) of the domain of the indicator	Conceptual definition (*meaning*) of the indicator (a benchmark for the qualitative dimension)
Logical	Logical specification of the query(-ies) needed to retrieve all the information (data) needed to calculate the indicator	*Data* definition: selection of the relevant information through the query
Functional	Mathematical expression of the indicator (to be applied to the results of the queries)	*Mathematical* definition: related to the selected method of calculation of the indicator (most relevant for the user: the user is interested in the value of the indicator!) Note that the method is outside the ontological domain
Qualitative	Ontological questions related to the meaningfulness of the indicator.	Definition of the criteria for the *assessment* of the obtained result (degree of meaningfulness of the indicator)

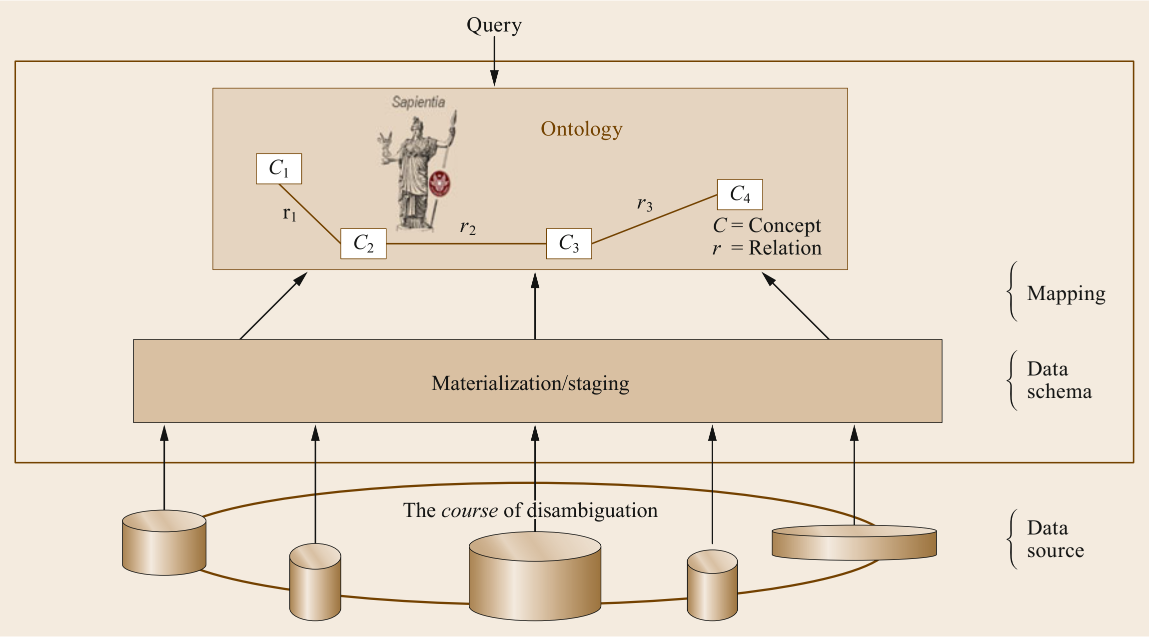


Fig. 15.12 Illustration of the materialization phase in an OBDI system

15.6 Conclusions

In this chapter, we introduced the main challenges in data integration for R&I. We discussed the two main existing approaches to data integration, namely procedural and declarative. We followed the latter approach and focused the subsequent analysis on the OBDI approach. The key idea of OBDI is to resort to a three-level architecture, constituted by the ontology, the sources, and the mapping between the two.

Daraio, *Lenzerini*, et al. [15.42] introduce the OBDI approach to coordinate, integrate and maintain the data needed for science, technology, and innovation policy and illustrate its potentials for specifying STI indicators and developing *science of science* policies. They outline the main advantages of OBDI with respect to the traditional silos-based approach to data integration, namely: *conceptual access to the data*, *re-usability*, *documentation and standardization*, *flexibility*, *extensibility*, and *opening of the system* (Table 15.3).

The three main advantages of OBDI for research and innovation analysis [15.33], which encompass and further expand those listed in Table 15.3 are: *openness*, interoperability, and data quality.

An OBDI approach may be an adequate platform/infrastructure to embrace and coordinate in an effective way (i. e., ensuring interoperability and high level of data quality standard), the many initiatives that are going on in research and innovation data collections.

Table 15.3 Main advantages of an OBDI approach over a traditional *silos*-based approach (after [15.42])

Advantage	Short description
Conceptual access to the data re-usability	Users can access the data by using the elements of the ontology.
Documentation and standardization	The mapping layer explicitly specifies the relationships between the domain concepts and the data sources. It is useful for documentation and standardization purposes.
Flexibility of the system	You do not have to merge and integrate all the data sources at once which could be extremely costly.
Extensibility of the system	You can incrementally add new data sources or new elements (ability to follow the incremental understanding of the domain) when they become available.
Opening of the system	Provide a conceptual framework that can be used as a common language by the community.

Figure 15.13 shows an outline of the main component of an open OBDI infrastructure.

An open STI data platform may encourage and support new research developments in the generation of new indicators carried out by scientists, which exploit the accessibility and transparency of data. In this way, there may be opportunities for the creation of new indicators beyond the short-term needs of policy-makers. The open-data framework, offers the possibility of full documentation on data and explicit articulation of logical linkages. It makes the traditional training and accreditation approach obsolete, in which users of indicators were dependent on the training provided by the owners of the data. Communities of users can, in fact, contribute to the improvement of the documentation and identify pitfalls and shortcomings of indicators. *Sapientia* and OBDI are two technologies to operationalize the consideration of data as infrastructural resources, as they are "shared means to many ends" that satisfy the three criteria of infrastructure resources [15.44]: a) non-rivalrous goods; 2) capital goods; and c) general-purpose inputs. *Sapientia* and OBDM could, indeed, be two *enabling technologies* to improve the exploitation of data for supporting growth and well-being, as proposed by OECD [15.45], and pushing towards the realization of an open science [15.46].

As we have showed in this chapter, the application of OBDI for the integration of data in R&I can be an interesting technology to further explore and exploit. However, it is important to point out that OBDM is not a *panacea* able to solve all the main challenges in data integration for R&I. Nevertheless, it can be a useful tool for *reasoning* over the assessment of research and innovation for different purposes and may lead to a more aware and careful specification of data and indicators useful in the evaluation process. *Sapientia* and its OBDI system may be at the heart of an open and collaborative platform around which to build a knowledge infrastructure for the assessment of research and its impacts.

Besides, the field of OBDI in computer science is far from being stable and consolidated. It is a dynamic and evolving field. *De Giacomo* et al. [15.14] discuss the main challenges related to OBDI that currently deserve investigation, namely querying rewriting optimization, meta-modeling, and meta-querying, non-relational data sources, OBDI methodology and tools, OBDI evolution, and going beyond data access. This is to say that the field of OBDI introduced in this chapter is a relatively new discipline to apply to interesting and complex issues, such as the evaluation of research and innovation, to corroborate the existing methodologies, and develop new and appropriate tools and solutions for data integration in these fields.

We conclude this chapter, recalling the final observation of *Daraio* and *Glänzel* [15.9]:

> One of the main Grand Challenges that remains to address is the exploitation of data availability, Information Technology and current state of the

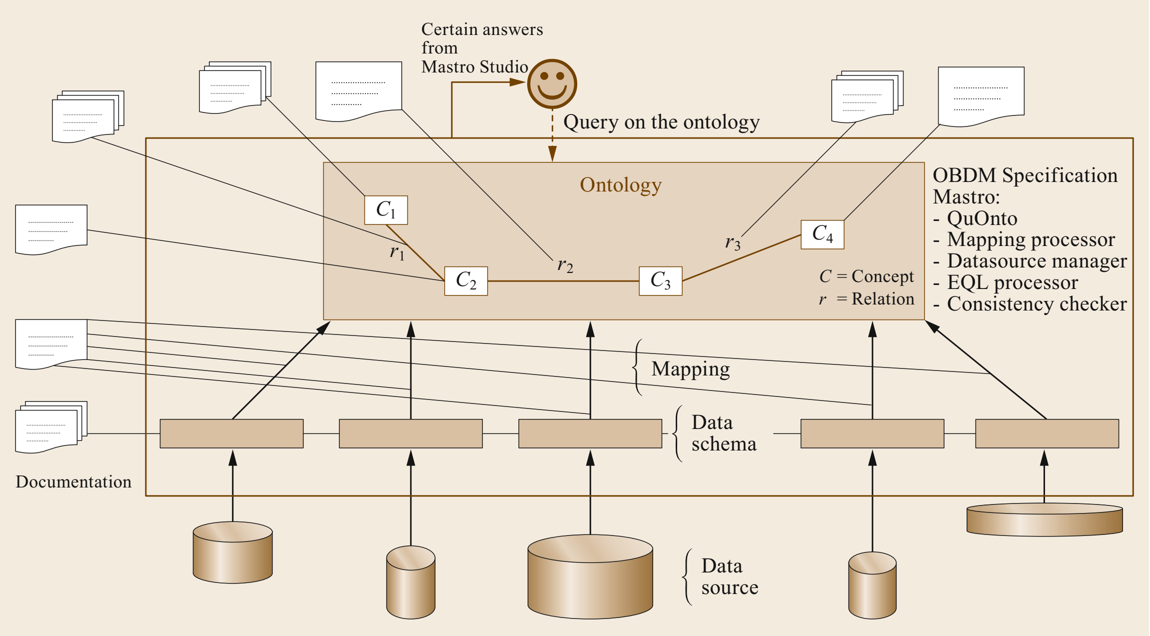


Fig. 15.13 Illustration of an open OBDI information system (after [15.33])

art in science and technology for the dynamical setting of standards in a data integration framework in use for multiple purposes.... Within this framework, to deal with this Grand Challenge, the interaction with stakeholders for ensuring an efficient and effective sustainable model is crucial. It depends also on the ability to successfully address, in a systematic way, the other problems highlighted above.

What would be needed to deal with this Grand Challenge and interact with stakeholders in order to develop and implement an OBDI approach for the integration of research information systems and the building of indicators upon them is a *long term investment* in the technology behind the OBDI and its related research. As recalled above, it is a relatively new discipline, which is worth further exploration and exploitation to address the existing challenges in data integration for R&I.

Acknowledgments. Financial support from the Project Sapienza Awards 2015 n. C26H15XNFS and the Project FILAS RU 2014-1186 is gratefully acknowledged.

References

15.1 J. Chen, Y. Chen, X. Du, C. Li, J. Lu, S. Zhao, X. Zhou: Big data challenge: a data management perspective, Front. Comput. Sci. **7**(2), 157–164 (2013)

15.2 H. Ekbia, M. Mattioli, I. Kouper, G. Arave, A. Ghazinejad, T. Bowman, V. Ratandeep Suri, A. Tsou, S. Weingart, C.R. Sugimoto: Big data, bigger dilemmas: A critical review, J. Assoc. Inf. Sci. Technol. **66**(8), 1523–1545 (2015)

15.3 C.L. Borgman: *Big Data, Little Data, No Data: Scholarship in the Networked World* (MIT Press, Cambridge 2015)

15.4 Z. Majkić: *Big Data Integration Theory, Texts in Computer Science* (Springer, Switzerland 2014)

15.5 X.L. Dong, D. Srivastava: Big data integration, Synth. Lect. Data Manag. **7**(1), 1–198 (2015)

15.6 M. Lenzerini: Data integration: A theoretical perspective. In: *Proc. 21st ACM-SIGMOD-SIGART Symp. Princ. Database Syst. PODS2002* (2002) pp. 233–246

15.7 C. Parent, S. Spaccapietra: Database integration: the key to data interoperability. In: *Advances in Object-Oriented Data Modeling*, ed. by M.P. Papazoglou, Z. Zari (MIT Press, Cambridge 2000) pp. 221–253

15.8 C. Daraio: A framework for the assessment of research and its impacts, J. Data Inf. Sci. **2**(4), 7–42 (2017)

15.9 C. Daraio, W. Glänzel: Grand challenges in data integration—state of the art and future perspectives: An introduction, Scientometrics **108**(1), 391–400 (2016)

15.10 OECD: *Quality Framework and Guidelines for OECD Statistical Activities* (OECD, Paris 2011)

15.11 W. Glänzel, S. Katz, H. Moed, U. Schoepflin: Preface, Scientometrics **35**(2), 165–166 (1996)

15.12 W. Glänzel, H. Willems: Towards standardisation, harmonisation and integration of data from heterogeneous sources for funding and evaluation purposes, Scientometrics **106**(2), 821–823 (2016)

15.13 W. Glänzel: The need for standards in bibliometric research and technology, Scientometrics **35**(2), 167–176 (1996)

15.14 G. De Giacomo, D. Lembo, M. Lenzerini, A. Poggi, R. Rosati: Using ontologies for semantic data integration. In: *A Comprehensive Guide Through the Italian Database Research Over the Last 25 Years, Studies in Big Data*, Vol. 31, ed. by S. Flesca, S. Greco, E. Masciari, D. Saccà (Springer, Cham 2018)

15.15 C. Daraio, M. Lenzerini, C. Leporelli, P. Naggar, E. Fusco, A. Balducci: Sapientia (the ontology of multidimensional research assessment) and OBDM (ontology based data management) as two key enabling technologies for the development of integrated data platforms for science, technology and innovation (STI). In: *OECD Blue Sky 2016, Ghent* (2016)

15.16 J.D. Ullman: Information integration using logical views. In: *Proc. Int. Conf. Database Theor., ICDT'97*, LNCS, Vol. 1186 (Springer, Berlin, Heidelberg 1997) pp. 19–40

15.17 A.Y. Levy, A.O. Mendelzon, Y. Sagiv, D. Srivastava: Answering queries using views. In: *Proc. 14th ACM-SIGMOD-SIGART Symp. Princ. Database Syst., PODS'95* (1995) pp. 95–104

15.18 A.Y. Halevy, A. Rajaraman, J. Ordille: Data integration: The teenage years. In: *Proc. 32nd Int. Conf. Very Large Data Bases, VLDB 2006* (2006) pp. 9–16

15.19 N.F. Noy, A. Doan, A.Y. Halevy: Semantic integration (editorial), AI Magazine **26**(1), 7 (2005)

15.20 D. Calvanese, G. De Giacomo, M. Lenzerini, R. Rosati, G. Vetere: DL-Lite: Practical reasoning for rich DLs. In: *Proc. Int. Workshop Descr. Log., DL2004*, CEUR, Vol. 104 (2004), http://ceur-ws.org

15.21 D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, R. Rosati: Tractable reasoning and efficient query answering in description logics: The DL-Lite family, J. Autom. Reason. **39**(3), 385–429 (2007)

15.22 A. Poggi, D. Lembo, D. Calvanese, G. De Giacomo, M. Lenzerini, R. Rosati: Linking data to ontologies. In: *J. Data Semant*, Vol. 4900 (Springer, Berlin, Heidelberg 2008) pp. 133–173

15.23 M. Lenzerini: Ontology-based data management. In: *Proc. 20th ACM Int. Conf. Inf. Knowl. Manag., CIKM'11* (2011) pp. 5–6

15.24 C. Daraio, M. Lenzerini, C. Leporelli, H.F. Moed, P. Naggar, A. Bonaccorsi, A. Bartolucci: Sapientia: the ontology of multi-dimensional research assessment. In: *Proc. 15th Int. Soc. Scientometr. Informetr. Conf., Istanbul*, ed. by A.A. Salah, Y. Tonta, A.A. Akdag Salah, C. Sugimoto, U. Al (Bogaziçi Univ. Printhouse, Turkey 2015) pp. 965–977

15.25 F. Baader, D. Calvanese, D. McGuinness, D. Nardi, P.F. Patel-Schneider (Eds.): *The Description Logic Handbook: Theory, Implementation and Applications*, 2nd edn. (Cambridge Univ. Press, Cambridge 2007)

15.26 T. Imielinski, W. Lipski Jr.: Incomplete information in relational databases, J. ACM **31**(4), 761–791 (1984)

15.27 S. Ceri, G. Gottlob, L. Tanca: *Logic Programming and Databases* (Springer, Berlin 1990)

15.28 R. Fagin, G.P. Kolaitis, R.J. Miller, L. Popa: Data exchange: Semantics and query answering, Theor. Comput. Sci. **336**(1), 89–124 (2005)

15.29 P.N. Edwards, S.J. Jackson, M.K. Chalmers, G.C. Bowker, C.L. Borgman, D. Ribes, M. Burton, S. Calvert: *Knowledge Infrastructures: Intellectual frameworks and research challenges* (Deep Blue, Ann Arbor 2013), http://hdl.net/2027.42/97552

15.30 N. Georgescu-Roegen: The economics of production, Am. Econ. Rev. **60**(2), 1–9 (1970)

15.31 N. Georgescu-Roegen: Process analysis and the neoclassical theory of production, Am. J. Agric. Econ. **54**(2), 279–294 (1972)

15.32 N. Georgescu-Roegen: Methods in economic science, J. Econ. Issues **13**(2), 317–328 (1979)

15.33 C. Daraio, M. Lenzerini, C. Leporelli, P. Naggar, A. Bonaccorsi, A. Bartolucci: The advantages of an ontology-based data management approach: Openness, interoperability and data quality, Scientometrics **108**(1), 441–455 (2016)

15.34 X. Li, J.D. Johnson: Evaluate IT investment opportunities using real options theory, Inf. Resour. Manag. J. **15**(3), 32–47 (2002)

15.35 C.Y. Baldwin, K. Clark: *Design Rules – The Power of Modularity* (MIT Press, Cambridge 2000)

15.36 D.L. Parnas: On the criteria to be used in decomposing systems into modules, Commun. ACM **15**(12), 1053–1058 (1972)

15.37 H.A. Simon: The architecture of complexity, Proc. Am. Philos. Soc. **106**, 467–482 (1962)

15.38 D. Lembo, D. Pantaleone, V. Santarelli, D.F. Savo: Easy OWL drawing with the graphol visual ontology language. In: *Proc. 15th Int. Conf. Princ. Knowl. Represent. Reason., KR2016* (2016) pp. 573–576

15.39 D. Lembo, D. Pantaleone, V. Santarelli, D.F. Savo: Eddy: A graphical editor for OWL 2 ontologies. In: *Proc. 25th Int. Jt. Conf. Artif. Intell., IJCAI* (2016) pp. 4252–4253

15.40 OECD: Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development. In: *The Measurement of Scientific, Technological and Innovation Activities* (OECD, Paris 2015), https://doi.org/10.1787/9789264239012-en

15.41 REF Research Excellence Framework: Panel Criteria and Working Methods. https://www.ref.ac.uk/2014/media/ref/content/pub/panelcriteriaandworkingmethods/01_12_1.pdf (2012)

15.42 C. Daraio, M. Lenzerini, C. Leporelli, F.H. Moed, P. Naggar, A. Bonaccorsi, A. Bartolucci: Data integration for research and innovation policy: An ontology-based data management approach, Scientometrics **106**(2), 857–871 (2016)

15.43 C. Daraio, A. Bonaccorsi: Beyond university rankings? Generating new indicators on universities by linking data in open platforms, J. Assoc. Inf. Sci. Technol. **68**, 508–529 (2016)

15.44 B.M. Frischmann: *Infrastructure: The Social Value of Shared Resources* (Oxford Univ. Press, New York 2012)

15.45 OECD: *Data-Driven Innovation Big Data for Growth and Well-Being* (OECD, Paris 2015)

15.46 OECD: Making Open Science a Reality. In: *OECD Science, Technology and Industry Policy Papers*, Vol. 25 (OECD, Paris 2015), https://doi.org/10.1787/5jrs2f963zs1-en

Maurizio Lenzerini

Department of Computer, Control, and Management Engineering
Sapienza University of Rome
Rome, Italy
lenzerini@diag.uniroma1.it

Maurizio Lenzerini is a Full Professor of Data Management and Semantic Technologies at the University of Rome La Sapienza. He has provided fundamental contributions to data management, knowledge representation and automated reasoning, and ontology-based data access and integration. He is the author of more than 350 publications. He is the recipient of two IBM Faculty Awards, a Fellow of EurAI, Fellow of the AAAI, and Fellow of the ACM.

Cinzia Daraio

Department of Computer, Control, and Management Engineering
Sapienza University of Rome
Rome, Italy
daraio@diag.uniroma1.it

Cinzia Daraio is an Associate Professor of Management Engineering at the University of Rome “La Sapienza”, where she teaches Productivity and Efficiency Analysis, Quantitative Models for Economic Analysis and Management and Economics and Business Organization. She is specialized in science and technology indicators, higher education microdata and methodological and empirical studies in productivity and efficiency analysis. She has authored more than 150 publications.

16. Synergy in Innovation Systems Measured as Redundancy in Triple Helix Relations

Loet Leydesdorff, Inga Ivanova, Martin Meyer

The Triple Helix (TH) of university–industry–government relations can first be considered as an institutional network. However, the correlations in the patterns of relations provide another topology: that of a vector space. Meanings are provided from positions in this latter topology and from the perspective of hindsight. Meanings can be shared, and sharing generates redundancy. Increasing redundancy provides new options and reduces uncertainty; reducing uncertainty improves the innovative climate, and the generation of options (redundancy) is crucial for innovation. The knowledge base provides an engine of the economy by evolving in terms of generating new options. The trade-off between the evolutionary generation of redundancy and the historical variation providing uncertainty can be measured as negative and positive information, respectively. In a number of studies of national systems of innovation (e.g., Sweden, Germany, Spain, China), this TH synergy indicator has been used to analyze regions and sectors in which uncertainty was significantly reduced. The quality of innovation systems can thus be quantified at different geographical scales and in terms of sectors such as high- and medium-tech manufacturing or knowledge-intensive services (KIS).

16.1 The Triple Helix Model of Innovations

The Triple Helix of university–industry–government relations emerged as a research program from a confluence of *Henry Etzkowitz*'s longer term interest in the entrepreneurial university [16.1–4] with *Loet Leydesdorff*'s interest in the evolutionary dynamics of science, technology, and innovation. Etzkowitz contributed a chapter entitled *Academic-Industry Relations: A Sociological Paradigm for Economic Development* to the book *Evolutionary Economics and Chaos Theory: New directions in technology studies* [16.5]. Leydesdorff argued in the Epilogue of this book that more than two interacting dynamics are needed for studying technology and innovation from an evolutionary perspective. Trajectories can be stabilized historically as a result of *mutual shaping* between two dynamics, but a third dynamic can be expected to disturb (destabilize) this tendency toward equilibrium and contribute to shaping a next-order (globalized instead of stabilized) regime. Different from an observable trajectory structuring behavior, a technological regime structures the expectations [16.6].

The sociologist *Simmel* noted already in 1902 that the transition from a dyad to a triad is fundamental to systems formation [16.7, 8]. A triad can be commu-

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_16

tative: are the friends of my friends also my friends? The order of the communications in a triad can be expected to generate asymmetries: two loops in one direction and one in the other may lead to a path different from that resulting from one loop in the first direction and two in the opposite. This system thus becomes path-dependent: one cannot go back without friction to a previous state, as in an equilibrium. An innovation system develops historically along trajectories. The triadic overlay has a dynamic different from the sum of the bilateral relations. It provides an emerging selection environment at the next-order level of an emerging regime.

Etzkowitz and *Leydesdorff* [16.9] considered this emerging operation as a *communication overlay*. *Ivanova* and *Leydesdorff* [16.10, 11] characterized the resulting communication system as a *fractal manifold*: the bilateral arrangements can be broken open (at all scales) by the third along each side of the triangle. A fractal manifold is scale-free because it develops endogenously in terms of reconstructions (which are needed because of the fractioning). The triads are nested at different levels and along different axes. In other words, a complex system can be expected to develop which is both horizontally and vertically differentiated [16.12].

The TH has hitherto focused on horizontal differentiation and integration among universities, industries, and governments as institutional spheres. In this chapter, we report on the further elaboration of the institutional model of university–industry–government relations into an evolutionary model of innovations as a vertical differentiation. We shall argue that the knowledge base evolves in terms of providing new options by making distinctions possible [16.13]. Increases (and decreases) in the number of options can be measured in terms of a trade-off between redundancy and uncertainty generation. We discuss the development of an instrument for the measurement of this balance in TH relations.

16.2 Institutional and Evolutionary TH Models

If sufficiently complex, institutional networks can carry the evolution of a knowledge base. However, the knowledge base develops with another dynamics on top of the institutional layer in terms of functions such as *supply*, *demand*, and *control* (Fig. 16.1).

These functions have to be specified at each level and sector. In a study of medical innovations, for example, *Nelson* et al. [16.15] distinguished among demand articulation for innovation in terms of diseases, supply in terms of new treatments, and control in terms of practical experiences and evaluations. *Petersen* et al. [16.14] measured these three dynamics in terms of medical subject headings (MeSH terms that are attributed by PubMed/MEDLINE). Branch *C: diseases* of this index can be considered as an articulation of demand, *D: drugs and chemicals* as supply, and *E: techniques and equipment* as conditioning the translations between supply and demand. Using the TH indicator—to be discussed below—windows of opportunities for new medical technologies were indicated.

The evolutionary and institutional dynamics are related in the events; for example, as coclassifications of publications by PubMed/MEDLINE. Unlike the TH model of university–industry–government *relations*, the evolutionary model includes both relations and nonrelations or, in other words, *correlations* among distributions of relations. The correlations span a vector space in which the relations and the carriers at the nodes occupy *positions*. In the case of three distributions, correlations between each two distributions can be spurious because of the third one. This latent dimension can contribute with either a plus or a minus sign. For example, university–industry relations in a nation or region may be excellent because of government policies or despite these policies.

In other words, each third helix can feedback or feedforward on the relations between the other two. From the perspective of this generalized TH model,

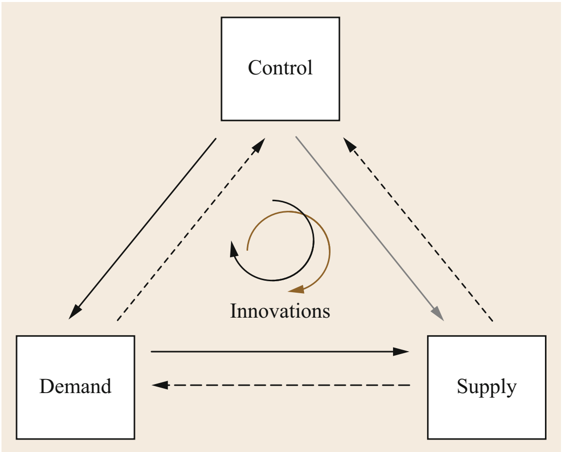


Fig. 16.1 The generalized Triple Helix model of innovations (after [16.14, Fig. 1, p. 667], with permission from Elsevier)

university–industry–government relations can be considered as the special case of focusing on institutional relations. From the evolutionary perspective, the analysis of relations is not a purpose but a means to study the potential synergy in new arrangements. The institutions and their relations develop historically and are therefore directly observable (*phenotypically*). However, functions can be specified as theoretically informed expectations (*genotypically*; [16.24]). The evolutionary TH model assumes that institutional arrangements evolve because of new options for:

(i) Knowledge production
(ii) Wealth generation, and
(iii) Regulation.

Table 16.1 summarizes the differences between the institutional and evolutionary model. The two models are intrinsically related as models for explaining emergence [16.25]. The institutional TH model focuses on relations [16.26, 27]. The relational interactions can be considered as first-order attributes to the nodes (in this case, institutions; e.g., *Etzkowitz* and *Zhou* [16.19]). Interactions among the relations in a triadic (or more-dimensional) configuration lead to a second-order dynamics among the first-order attributes (Fig. 16.2).

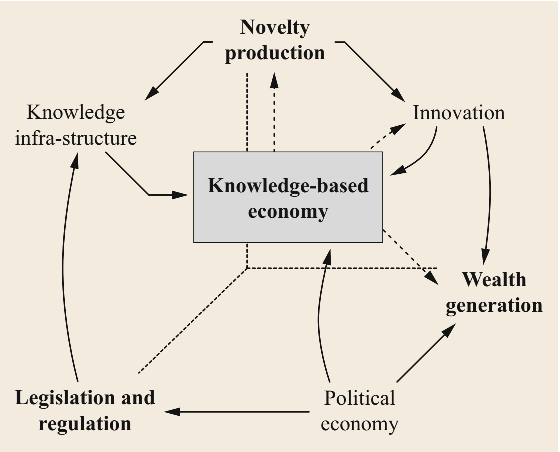
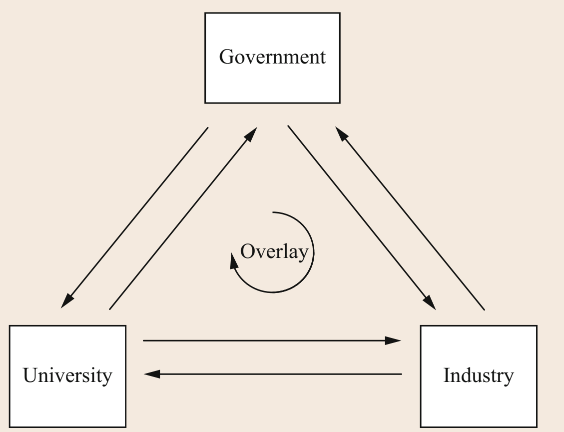
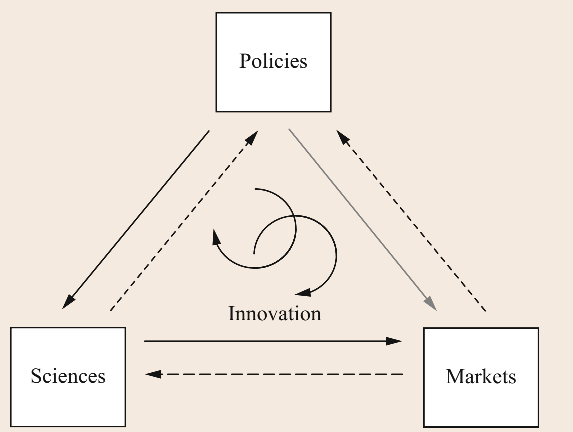


Fig. 16.2 Interactions among the first-order interactions generate a knowledge-based economy as a next-order system (after [16.28, p. 379], with permission from Wiley)

Table 16.1 Summary of the differences between the institutional and evolutionary TH models

<table>
<tr><th>Institutional TH model of university–industry relations</th><th>The evolution of TH relations in interactions among markets, sciences, and policies</th></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>

- University–industry–government relations
- (Inter-)institutional
- Entrepreneurship (*agents*)
- Network analysis; graphs
- Historical cases (*phenotypes*)
- Inductive:
  - *Best practices*; comparative case studies [16.16]
  - Bottom-up [16.17]
  - Policy analysis [16.18–20]

[16.9, 21, 22]</td><td>

- Correlations among social coordination mechanisms
- Evolutionary modeling of innovations (*constructs*)
- In the vector space:
  - TH synergy indicator
  - Redundancy (overlap) as a source of innovations

[16.10, 11, 23]</td></tr>
</table>

Whereas the functions of wealth generation and governance have been central to the analysis of political economy, the study of the knowledge-based economy includes the additional dynamics of innovations ([16.29]; see Fig. 16.2; [16.30, p. 186f.]; [16.31]). The trilateral interactions among the bilateral ones generate a feedback on the constituent helices and their bilateral relations *by providing an emerging selection environment*.

The institutional dynamics lead to path dependencies along historical trajectories. Between the historical variation and the selection mechanisms operating evolutionarily one can expect a nonlinear dynamics of creative destruction, agglomeration [16.32, 33], and reconstructions on the basis of competitive advantages [16.34]. The institutional restructuring is one of the relevant subdynamics of the complex dynamics of societal innovation and entrepreneurship. The relations and arrangements among institutions furthermore are sensitive to policy interventions. The objective, however, is the generation of synergy. The criterion of generating synergy in configurations of relations can provide a frame of reference for institutional reform since the lower the resulting uncertainty, the *smarter* specializations in terms of options can be.

In summary, the TH cannot be considered as a single method or model; it is a theme that binds together the transition of political economy into a knowledge-based economy as a macrodevelopment with the study of transitions at micro- and mesolevels based on and leading to knowledge-based innovations. The study of knowledge-based economic developments can be pursued at both institutional and functional levels. The institutional dynamics provide the social embedding to the evolving systems [16.35]. The model can be considered institutional insofar as the explanation is in terms of networked relations and interactions among institutional agents [16.26, 36]. The model is *neo*-evolutionary since more than a single selection is assumed and the selection environments can change and interact [16.37, 38]. The relational networks generate and retain the evolutionary dynamics [16.19, 25].

Knowledge-based innovation systems can be studied at macro-, meso-, or microlevels. Using the TH indicator, one is able to decompose the macro in terms of micro- and mesolevels. Thus, the issue of whether systems of innovation are national, regional, sectorial, etc., [16.39, 40], can be addressed empirically. A systemic development can be distinguished from nonsystemic (e. g., incidental) covariation. One can ask, for example, how much synergy is indicated at regional or national levels. Furthermore, one can quantify how much cross-border synergy the national level adds to the sum of the regional systems of innovation.

Note that the TH indicator provides a specific methodology which does not have to be used in a TH study. Much depends on the research questions and the kinds of data available. Crucial for the TH theme, in our opinion, is the extension of the economic and political analysis with attention to cognitive structuration [16.41]. Let us first specify the transition towards a knowledge-based economy in the macrohistorical context, and then proceed to discuss quantitative studies using contemporary data.

16.2.1 The Emergence of a Knowledge-Based Economy

In his time, Marx witnessed the prelude to the emergence of a knowledge-based economy. He was thoroughly aware of this. After studying in the British Library for almost a decade, in 1857 Marx published his rough draft of *Capital* under the title *Grundrisse: Foundations of the Critique of Political Economy* [16.42]. In this study, Marx considered the possibility of a knowledge-based economy as an alternative to the political economy that he criticized for ideologically accepting the extraction of wealth from labor, and thus condoning exploitation. As he put it (at p. 706):

> Nature builds no machines, no locomotives, railways, electric telegraphs, self-acting mules etc. These are products of human industry; natural material transformed into organs of the human will over nature, or of human participation in nature. They are *organs of the human brain, created by the human hand;* the power of knowledge, objectified. The development of fixed capital indicates to what degree general social knowledge has become a *direct force of production,* and to what degree, hence, the conditions of the process of social life itself have come under the control of the general intellect and been transformed in accordance with it. To what degree the powers of social production have been produced, not only in the form of knowledge, but also as immediate organs of social practice, of the real life process.

Note that Marx specified an indicator of this transition: *the development of fixed capital*. He discussed its operationalization at length (in the *Grundrisse*) and set himself the task to study the possibility that science and technology had become greater sources of societal wealth than labor. A model with two independent variables was not available in his time.

After another ten years of study, *Marx* [16.43] concluded in *Capital* that the main contradiction at the time remained the one between capital and labor. In the footnotes as a subtext (e. g., p. 393, note 89), however, Marx repeats that "the technology shows us the active relation of the human kind to nature, the immediate production process of our lives ..." If technology could enable us to free man from work sufficiently, the nature of capitalism would change, since "*the basis of this mode of production falls away*" (p. 709; italics in the original). In other words, Marx envisaged a regime change that is different from and an alternative for the communist revolution.

William Henry Perkin's research on dyestuffs in England during the late 1850s, for example, developed into an industry in Germany [16.44] [16.45, pp. 161f.] [16.18, p. 25]). However, *Noble* [16.46, p. 7] argued that "the major breakthroughs, technically speaking, came in the 1870s." He dated what he calls "the wedding of the sciences to the useful arts" as the period between 1880 and 1920. *Braverman* [16.45] introduced the concept of a *scientific-technical revolution* for indicating this same period when he formulated the regime change as follows (pp. 166f.):

> The scientific-technical revolution ... cannot be understood in terms of specific innovations—as is the case of the Industrial Revolution, which may be adequately characterized by a handful of key inventions—but must be understood rather in its totality as a mode of production into which science and exhaustive engineering investigations have been integrated as part of ordinary functioning. The key innovation is not to be found in chemistry, electronics, automatic machinery, aeronautics, atomic physics, or any of the products of these science-technologies, but rather in the transformation of science itself into capital.

The incorporation of science and technology into the production process makes the system evolve with a different dynamics [16.29]. Whereas both markets and political institutions can be considered as equilibrium-seeking [16.47], the nonequilibrium dynamics of the social production of knowledge makes evolution theory relevant to the analysis of innovation systems [16.48–51]. After WW II, the new field of *evolutionary economics* gradually emerged as central to innovation studies [16.52, 53]. However, it took until the 1980s before the debate about the knowledge-based economy and its institutional conditions became salient. Before that time, the confrontation between liberal and communist models of political economy dominated the Cold War. With the demise of the Soviet Union (1991) and the opening of China after 1989, this debate about political economy lost its prominence. The study of the emergence of a knowledge-based economy became urgent.

Using *Friedrich List*'s [16.54] model of national systems of political economy, *Freeman* [16.55] first proposed the model of *national systems of innovation* after studying Japan from a West-European perspective [16.56–58]. *Freeman* and *Perez* [16.35] further developed a macromodel of business cycles which updates Marx's dialectics of production relations and production forces. Using historical examples, these authors argue that long-term cycles (*technoeconomic paradigms*) are generated by *key factors* (such as oil in the previous cycle, or information in the current one) that can rapidly become abundant and thus cheaper. The structural crises between the new paradigm and existing institutions and industries call for adjustments. National innovation systems compete in terms of institutional reforms. The key factors which trigger next cycles, however, remained exogenous in this model, since the dynamics generating the *key factors* were not specified.

From a somewhat different perspective, *Nelson* and *Winter* [16.38, 59] called for evolutionary models of technological innovation that would endogenize the technological dimension. How is the knowledge base generated within the system? *Rosenberg* [16.60], for example, proposed to study selection in terms of focusing devices and inducement mechanisms. Under the condition of war, for example, national governments can be expected to invest in military technologies. Problems with the measurement of the knowledge base, however, seemed prohibitive in opening the black box further than in terms of historical descriptions [16.61, 62] or *history-friendly* models [16.50, 63]. How can one proceed from case descriptions and historical *phenotypes* to the specification of the evolutionary dynamics [16.52, 64]?

Within this program of studies, the issue of measurement became increasingly important. In his Presidential address to the American Economic Association, *Griliches* [16.65, p. 14] mentions the problem of measurement as a main constraint in research:

> After decades of discussion we are not even close to a professional agreement on how to define and measure the output of banking, insurance, or the stock market (see *Griliches* [16.66]). Similar difficulties arise in conceptualizing the output of health services, lawyers, and other consultants, or the capital stock of R&D. While the tasks are difficult, progress has been made on such topics.

How can one measure innovations and innovation systems?

For decades, *Freeman* and *Pavitt* curated a database of innovations at the Science Policy Research Unit (SPRU) of the University of Sussex [16.67]. In collaboration with Eurostat, in 1992 the OECD (Organization for Economic Co-operation and Development) developed the so-called *Oslo Manual* entitled *Guidelines for Collecting and Interpreting Innovation Data* [16.68]. The harmonization of national statistics is a first condition for making it possible to compare among national systems of innovation in terms of their strengths and weaknesses and perhaps to formulate best practices. However, neither innovation survey data nor patent data can be integrated easily—i. e., without additional assumptions—into the measurement of *national systems of innovation*. Patents, for example, are indicators of invention, and inventions are only proxies of innovation [16.65, 69, 70].

The focus on the national level was criticized by authors favoring regional perspectives [16.39] and resounded with the European Union's perspective on transnational and inter-regional innovation systems. The knowledge-based economy provided a metaphor which leaves the systems level open [16.40]. The economy is not defined institutionally, but functionally. At an expert meeting of the OECD in 1994 about developing indicators for the knowledge-based economy [16.71], however, *Carter* [16.72] warned that the measurement of knowledge had remained an unsolved problem [16.73]. *Andersen* [16.52] raised the question of *what is evolving?* in a knowledge-based economy as studied in *evolutionary economics*. Problems of operationalization and measurement thus came to the forefront.

During the second half of the 1990s, the OECD hosted a program about *the knowledge-based economy* [16.71, 74]. In an evaluation, *David* and *Foray* [16.75] cautioned that the terminology—knowledge-based economy—"marks a break in the continuity with earlier periods, more a 'sea-change' than a sharp discontinuity" [16.75, p. 9]. The authors noted that transformations can be analyzed at a number of different levels, and argued that *knowledge* and *information* should be distinguished more carefully by analyzing the development of a knowledge-based economy in terms of codification processes [16.76, 77].

Codification is a communication-theoretical problem: information can be provided with meaning, and specific meanings can further be codified as knowledge. The dynamics of the codification of information into knowledge are very different from the dynamics of government at different levels or the dynamics of markets in industrial sectors. The construction of knowledge-based systems is bottom-up, whereas retention of economic wealth from knowledge requires the downward arrow of control. How can the constructed advantages [16.34] be used for innovation? The theoretical challenge is to combine the perspective of codification with evolutionary and systemic perspectives [16.78].

16.3 The Operationalization of the Triple Helix

The Triple Helix model takes the challenges thus articulated, as starting points for further analysis and theorizing, but with the goal of operationalization and measurement. *David* and *Foray*'s [16.75] "break in the continuity with earlier periods" is appreciated as the new role and the transformative dynamics of the social organization of knowledge. This adds a third structural dynamic to the economy and to the political system regulating the economy. The dynamics of codification operate in terms of trajectories and regimes [16.6]. A regime has one more degree of freedom than a trajectory. Whereas trajectories are shaped in a landscape [16.63, 79], the knowledge-based regime is hypergeometrical. It can also be considered as the *genotype* of the observable *phenotypes* along trajectories.

The genotype of the technoeconomic evolution is not given like the DNA in biological systems. This *non-biological genotype* remains a construct that is open to partial reconstruction as knowledge is further developed [16.24]. In other words, *Andersen*'s [16.64] question of *what is evolving?* can now be answered: knowledge is a *cognitive* construct that evolves by generating new options. This evolving construct is socially retained and embedded along historical trajectories. The latter, however, are *social* constructs. Both dynamics are enacted and interact in events and actions generating and breaking relations.

Unlike the positive sciences which study domains that can be defined empirically, studies of sociocognitive systems develop cognitive means to study knowledge-based developments as their empirical domain. This raises questions of reflexivity [16.80–82]. The cognitive dimension cannot be observed as given naturalistically; it needs to be specified as a construct. The specification is performative in that it changes the expectations. (Expectations can be tested against observations!) Because of this constructivist constraint in the study of knowledge-based systems, the analysis remains at the edge of philosophy and develops what has been called an *empirical* [16.83, 84] or *concrete* philosophy of science [16.85, p. 159], [16.86]. This philosophy of science is concrete, since its status is one of hypothetical knowledge in need of the observation and testing of the contexts in which it is generated and which it transforms.

Do the observable instantiations inform us about the evolving system(s) by enabling us to improve our expectations? The measurement of science is nowadays further developed in the scientometric tradition. However, scientometricians are confronted with the same reflexivity problem—albeit from a different perspective—when the question is asked: what do the indicators indicate? When measuring, for example, the spectacular increases in international coauthorship relations in recent decades [16.87–89], does one indicate the growth of knowledge or only institutional expansion [16.90]? How do social relations reflect or perhaps interfere with cognitive constraints and opportunities? Can one infer from the one to the other dimension?

In an attempt to bridge the gap between qualitative theorizing and quantitative methods, *Callon* et al. [16.91] proposed juxtaposing the social, technical, and cognitive dimensions by considering networks of science and technology as heterogeneous. From this perspective, different units of analysis such as texts, authors, and cognitions are assembled in networks with a priori equal status (as actants). On the basis of the semiotic tradition, these actants are considered representations in a network whose dynamics can be mapped using cowords or other symbolic references, [16.92, 93].

Alternatively, one can distinguish among dimensions and relate networks of different character (e.g., [16.95, 96]. For example, one can use more-dimensional (n-mode) networks. Using Fig. 16.3, one

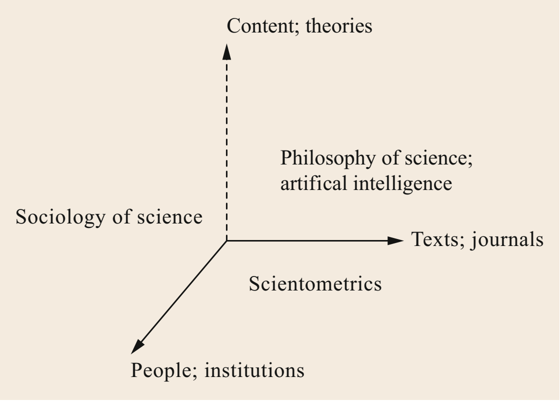


Fig. 16.3 Three main dimensions of science. Source: [16.94, p. 3]

of us proposed to distinguish in science studies more fundamentally among texts, people, and cognitions as three different units of analysis which cannot be reduced to one another. One can attribute texts to authors or vice versa, but the variation among the authors is different from the varieties among the texts. One can expect different—albeit interacting—dynamics along the three axes of cognitive content, social agency, and textual structures. Texts, for example, can be aggregated into journals or archives, whereas agents can be organized into institutions or groups. The dynamics of knowledge include, for example, the validation of new knowledge claims [16.97].

The three analytically distinguishable dynamics (in this case, social, cognitive, and textual) can also be considered as selection mechanisms operating upon one another. In a coevolutionary model, the variation-generating dimension (*helix*) can act as the selection mechanism at a next moment. Each two selection mechanisms operating upon each other can lead to *mutual shaping* along trajectories. A third selection environment, however, makes the system hyperselective [16.98]; one adds a *virtual order* or, in other words, "an absent set of differences, temporarily 'present' only in their instantiations, in the constituting moments of social systems" [16.41, p. 64]. One can expect skewed distributions in the outcome (e.g., scientometric distributions) because of selections operating upon one another [16.99].

In other words, we submit that the TH model can be appreciated in different topologies at the same time: the topology of the network of relations, and a vector space spanned by correlations providing structure in the background. In the latter, the zeros (nonrelations) are as important as the ones [16.100, pp. 3–9]. Structures operate deterministically at each moment of time, but over time the selection mechanisms operate upon one another, and thus structures are also at variance. Both relations and nonrelations can be considered as events that are selectively provided with meaning in the vector space. Meanings provided from different perspectives can be shared to differing extents.

The sharing of meanings generates overlap and redundancy, while the communications in terms of relations continuously generate variation containing (Shannon-type) information. Redundancy and information add up to the maximum entropy at each moment of time. However, the mechanisms of generating redundancy and information are different: information is generated historically, while redundancy is specified discursively in the knowledge base. The generation of redundancy—and not the generation of information—can make the system increasingly knowledge-based.

16.4 The Generation of Redundancy

Redundancy R was defined by *Shannon* [16.101] as follows

$$R = \frac{H_{\max} - H_{\mathrm{s}}}{H_{\max}} = 1 - \frac{H_{\mathrm{s}}}{H_{\max}} \,. \tag{16.1}$$

The maximum information content of a system ($H_{\max}$) is equal to the logarithm of the number of possible states N: i. e., $H_{\max} = \log(N)$. Equation (16.1) specifies that $H_{\max}$ is composed of two components: the system states hitherto realized ($H_{\mathrm{s}} = -\sum_i p_i \log(p_i)$) and the states which are possible given the definition of the system, but which were not (yet) realized: $H_{\max} - H_{\mathrm{s}}$.

For an innovation system, the number of options still available—that is, the redundancy—may be more important than the past record of already realized options; particularly when a system runs out of options. When additional redundancy is generated, the relative uncertainty $H_{\mathrm{s}}/H_{\max}$ decreases (ceteris paribus) because $H_{\max}$ increases. The exploration of new options (e. g., diversification) becomes less risky under the condition of less uncertainty [16.102]. The generation of redundancy, in other words, can be expected to improve the climate for entrepreneurship and innovation.

Shannon [16.101] deliberately abstained from the further specification of redundancy in terms of loops. From his perspective, redundancy and coding are needed for error-correction in the transmission (as "excess information"; [16.101]). Error-correction, however, assumes a norm and thus a social system. We note that meaning is provided from the perspective of hindsight and therefore against the axis of time. Insofar as meaning processing requires relationship and communication, Shannon-type information is generated, but at another level—that is, in the relational network space. The relational network can be rewritten as a matrix which can be analyzed in terms of eigenvectors. The focus on this vector space provides a different perspective on the same information contained in the distributions of historical events (e. g., relations).

Since *Shannon* [16.101] defined information as probabilistic entropy. (Shannon used H in Gibb's formulation of the entropy ($S = k_{\mathrm{B}}H$); k_{B} is the Boltzmann constant which provides the thermodynamic entropy S with the dimensionality Joule/Kelvin. However, H is dimensionless.) Because of the second law of thermodynamics, the development of information can only generate increasing entropy [16.103, 104]. The generation of redundancy, however, can be lead to either entropy increases or decreases depending on the feedback and feedforward loops in the meaning processing as different from information processing. Feedback and feedforward loops can be expected to propel information and meaning in clockwise or counter-clockwise cycles (Fig. 16.1), that is, with potentially opposite signs [16.10, 11, 105]. The relative information content of a message ($H_{\mathrm{s}}/H_{\max}$) can be enlarged or reduced by adding or constraining redundancy.

In other words, options other than those already realized are added or removed by mechanisms different from the second law. However, the number of options in a social system can increase much faster than their realizations. In a model, for example, the realizations are considered as special cases among possible states. As the models are refined, more distinctions and therefore options are made available. A knowledge-based economy seeks to exploit the increases of redundancy as a source of wealth.

16.5 The Triple Helix Indicator of Mutual Redundancy

How can redundancy be generated in relations among communication systems? Figure 16.4 visualizes a relationship between two sets as an overlap.

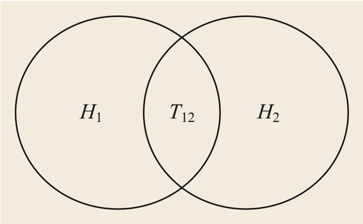


Fig. 16.4 Mutual information between two sets of messages

The formula for mutual information (or transmission T) follows according to the rules of set theory

$$H_{11} = H_1 + H_2 - T_{12} \,. \tag{16.2}$$

One subtracts the overlap (T_{12}) from the summation because otherwise one would count mutual information twice. However, if one sums the two sets as whole sets—accepting and including redundancy as surplus information—one obtains

$$\Upsilon_{12} = H_1 + H_2 + T_{12} \,. \tag{16.3}$$

Redundancy adds to the information content as "excess information" [16.101]. Comparing (16.2) and (16.3), mutual redundancy $R_{12} = -T_{12}$. Whereas T_{12} is Shannon-type information and therefore positive, R_{12} consequentially is expressed in terms of negative bits of information.

(16.2) can be rewritten in a more general format as follows

$$T_{12} = \sum_{i=1}^{n=2} H(x_i) - H(x_1, x_2) \geq 0 \,. \tag{16.4}$$

For more than two dimensions one can inductively generalize (16.4) as follows

$$\sum_{i=1}^{n=3} H(x_i) - H(x_1, x_2, x_3) = \sum_{ij}^{3} T_{ij} - T_{123} \tag{16.5}$$

$$\begin{aligned} &\ldots \\ &\sum_{i=1}^{n} H(x_i) - H(x_1, \ldots, x_n) \\ &\quad = \sum_{ij}^{\binom{n}{2}} T_{ij} - \sum_{ijk}^{\binom{n}{3}} T_{ijk} + \sum_{ijkl}^{\binom{n}{4}} T_{ijkl} - \ldots \\ &\qquad + (-1)^n \sum_{ijkl\ldots(n)}^{\binom{n}{n}} T_{ijkl\ldots(n)} \,. \end{aligned} \tag{16.6}$$

The left-side terms of (16.5) and (16.6) are positive because of the subadditivity of entropy (16.4). The alternation of the sign for mutual information in n dimensions (16.6)

$$(-1)^n \sum_{ijkl\ldots(n)}^{\binom{n}{n}} T_{ijkl\ldots(n)}$$

is an analytical consequence of this subadditivity. Taken apart, T_{ijk} and next-order terms can no longer be considered as Shannon-type information because of the sign changes [16.103, 104]. With the opposite sign, however, T_{ijk} can be considered as a measure of mutual redundancy. For n dimensions, the mutual redundancy R_n is

$$\begin{aligned} R_n &= -[(-1)^n T_{1234\ldots n}] \\ &= -\left[\sum_{i=1}^{n} H(x_i) - H(x_1, \ldots, x_n) \right] \\ &\quad + \left[\sum_{ij}^{\binom{n}{2}} T_{ij} - \sum_{ijk}^{\binom{n}{3}} T_{ijk} + \sum_{ijkl}^{\binom{n}{4}} T_{ijkl} - \ldots \right. \\ &\qquad \left. + (-1)^{1+n} \sum_{ijkl\ldots(n-1)}^{\binom{n}{n-1}} T_{ijkl\ldots(n-1)} \right] . \end{aligned} \tag{16.7}$$

R_n can be positive or negative: the first term on the right side of (16.7) is necessarily negative (because of the minus sign); but the second term is positive entropy in a set of relations. The outcome is balanced and therefore empirical. The more negative the sum, the more options are generated.

In other words, mutual redundancy is generated in a trade-off between selective structures and variable configurations of relations. A configuration can also be reorganized; for example, in terms of developing new institutional arrangements. The minimalization of the second (positive) term in (16.6) provides us with a criterion for the evaluation of changes in the relations. (We have hitherto not further elaborated this criterion.) The positive term is historically contingent, whereas the negative terms reflect the structure(s) in the system. As noted, these structures are not given naturally, but are (re)constructed. The technocultural evolution based on distinguishing [16.13] thus transforms the historical developments.

As new options are made available, the domain of what *Kauffman* [16.106] called *adjacent others*—diversification options at the border between historically realized and possible, as yet unrealized states—is changed. The shaping of new relations and loops changes the phase space first along historical trajectories. However, possibly unintended loops may emerge which feedback or feedforward on existing loops. Resonances among the loops can trigger a next-order cycle of redundancy generation, such as a change in a technological regime [16.107]. A change at the regime level implies a redefinition of the selection mechanisms in the vector space since another dimension is added. What *demand*, *supply*, and *control* mean may have changed after such a transition. For example, the demand for innovation in horse shoes changed after the introduction of the automobile. Although the automobile first emerged necessarily along a trajectory, the car system followed as a regime with many feedback loops. Feedback loops stabilize the system, whereas feedforward ones destabilize; but they enhance globalization beyond the boundaries currently given.

In summary, the information-theoretical perspective provides us with a model of technoeconomic evolution beyond the measurement instrument of the TH indicator. The regime level adds another selection environment reorganizing the trajectories [16.6, 108]. This selection environment—a communication field or overlay—emerges first as a second-order interaction term among bilateral relations (Fig. 16.2), but then becomes analytically different—as a knowledge base—from the selection environments from whose interactions and overlaps it emerged.

16.6 The Measurement

The TH indicator was first developed in the context of the institutional TH model as a quantification of the balance between bi- and trilateral relations among universities, industries, and governments [16.109, 110], [16.111, pp. 143 ff.]). The indicator can be derived using the Shannon formulas [16.112, 113] as

$$T_{123} = H_1 + H_2 + H_3 - H_{12} - H_{13} - H_{23} + H_{123} \,. \quad (16.8)$$

As noted above, T_{123} is not a Shannon measure since it can be negative. (The Shannon measure with a positive sign is $\sum_{ij}^{3} T_{ij} - T_{123} \geq 0$; (16.5) above). In the three-dimensional case, mutual redundancy $R_{123} = T_{123}$. In the two-dimensional case, however, $R_{12} = -T_{12}$. R measures mutual redundancy in a configuration of relations under study.

Figure 16.5, for example, is based on using all publications in the science citation index (SCI) with at least one South Korean address in the byline during the period 1973–2006. These publications carry 190 196 Korean addresses which were manually evaluated and then analyzed in terms of university–industry–government coauthorship relations using the TH indicator R_{123}. The figure shows the development of the interactions: whereas initially the system was state-controlled, the dictatorship regime relaxed gradually during the 1970s. This tendency was strengthened during the period of democratization in the 1980s. After the status of a more advanced economy is reached, the pendulum in the balance between uncertainty and redundancy generation swings back when Korea enters increasingly the world market, leading to full OECD membership in 1996. The internationalization of the research system uncouples from the national system of publications, and mutual redundancy thus decreases in absolute value (or, in other words, becomes less negative). Communication becomes more efficient or, in other words, less redundant.

On the basis of SCI data for the year 2011, Fig. 16.6 shows the strong integration at the national level in the

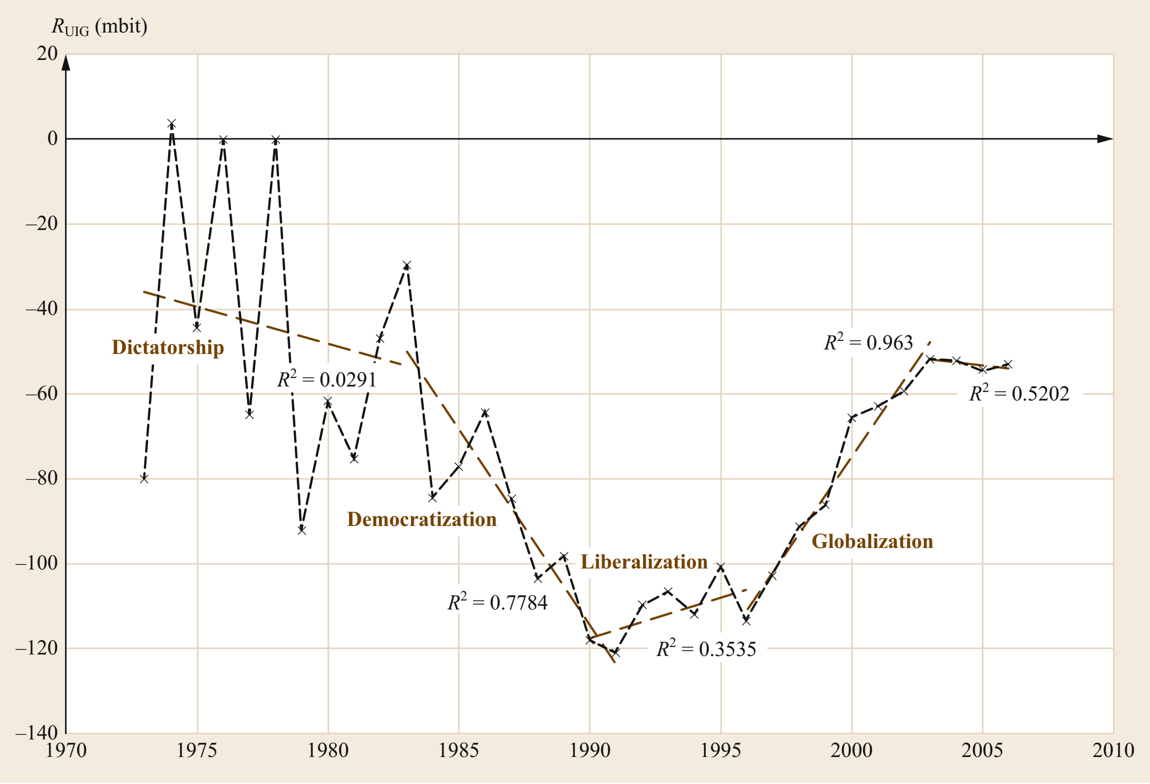


Fig. 16.5 The development of mutual redundancy in South Korean university–industry–government coauthorship relations during the dictatorship, the periods of democratization, liberalization, and globalization, respectively. (Adapted from [16.110, p. 645])

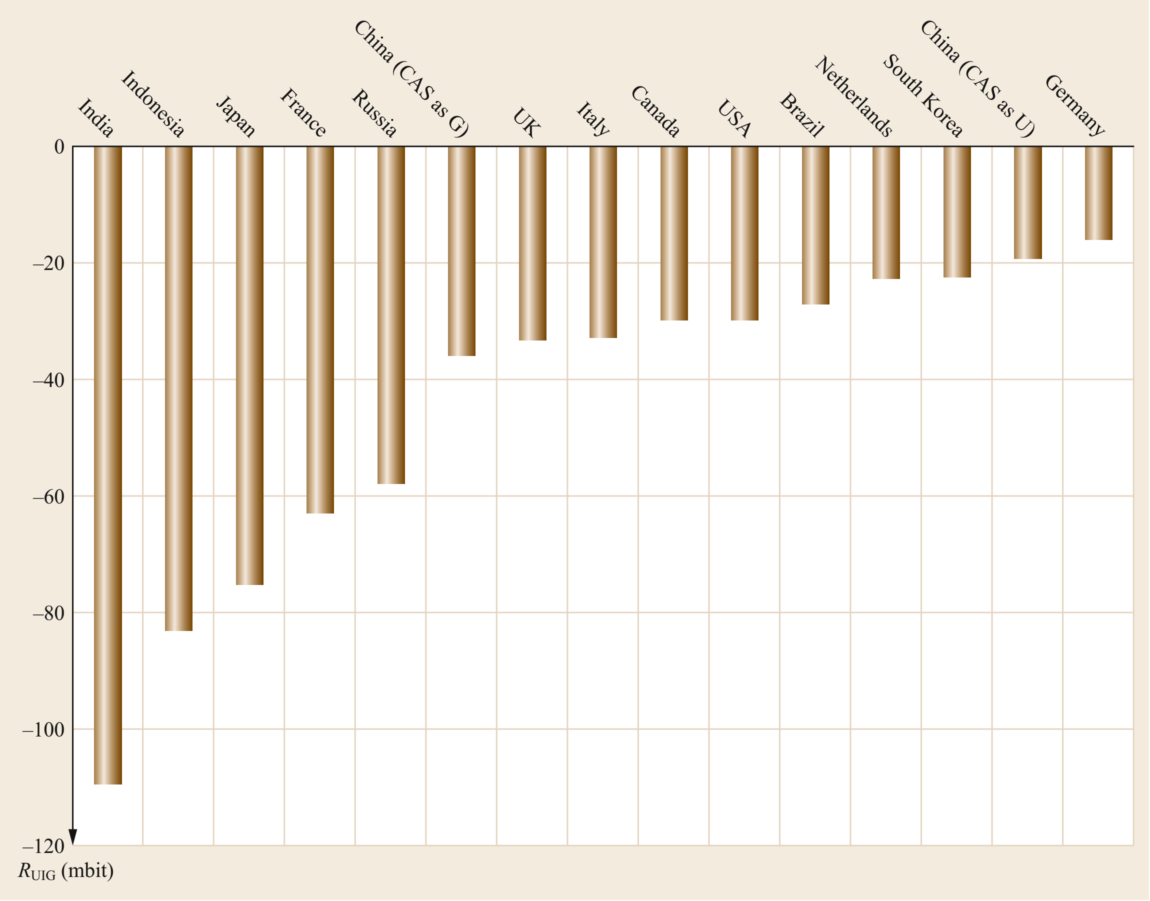


Fig. 16.6 University–industry–government coauthorship relations in 2011, evaluated in terms of mutual redundancy. Source: [16.114]. N of publications = 1 195 494 retrieved from the *Science Citation Index,* using an approximation developed by *Leydesdorff* [16.109, p. 458]; cf. *Park* et al. [16.115, 13 ff.]

case of some Asian countries (India, Indonesia, and Japan), whereas OECD member states (e. g., Germany and South Korea) are oriented more globally. The Chinese data provide us with an opportunity to consider publications of the Chinese Academy of Science (CAS) as either university or government.

The CAS is gradually making this transition [16.116]. When CAS publications are considered as university publications, the Chinese system can be compared with South Korea and Germany in terms of its local (national) versus global orientation. As government publications, however, CAS firmly anchors the Chinese publication system at the national level. The different patterns of TH configurations in developed versus developing nations have been further investigated by *Choi* et al. [16.117, 118].

16.7 Measuring the Knowledge Base of Innovation Systems

In studies focusing on university–industry–government relations, one can count seven instances (U, I, G, UI, UG, IG, and UIG) and then evaluate the combinations in the following three-dimensional cube of information (Fig. 16.7).

Note that the eighth option $\{U = 0; I = 0; G = 0\}$ is not counted in this (relational) model. Along each of the axes, however, one can refine the measurement. Instead of *university*, for example, one can distinguish among disciplines in terms of departments and faculties, given

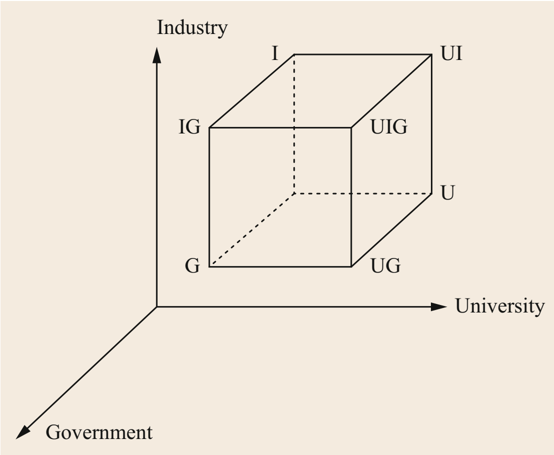


Fig. 16.7 University–industry–government relations and a three-dimensional vector space

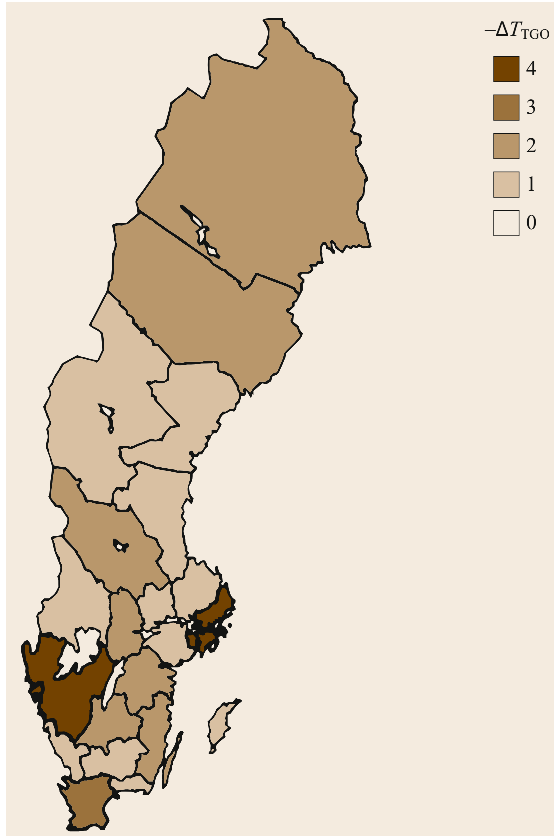


Fig. 16.8 Contributions to redundancy at the level of 21 Swedish counties (NUTS-3)

that university–industry relations are very different in biomedicine, engineering, or the social sciences. Similarly, industry can be differentiated among sectors (e. g., medium- and high-tech) and the dimension of government can be made more precise as national, regional, city, etc.

Using *Storper*'s [16.26] metaphor of a "holy trinity of technology, territory, and organization," one can organize firms in terms of technological classes, geographical addresses, and organizational size, and study the interactions among these three dimensions. Which regions or sectors contribute most to the generation of redundancy? In the case of Sweden, for example, the complete set of (micro) firm data for Sweden at Statistics Sweden was $N = 1\,187\,421$ in November 2011. This Swedish data contains address information in terms of 290 units at the lowest level of municipalities (NUTS5), a technology classification into 21 classes—concordant with the NACE classification of the OECD/Eurostat—and nine classes of numbers of employees which allow us to distinguish between small, medium-sized, and large companies using the TH calculator available at http://www.leydesdorff.net/software/th4 [16.119].

Figure 16.8 shows the results for the 21 counties in Sweden at the level NUTS-3. As could be expected, mutual redundancy is highest (in absolute value) for Stockholm (−3.49 mbit), Västre Gotalands län (−2.91 mbit), and Skåne (−2.31 mbit). These three counties host the major universities and dominate the picture within the nation; together they account for $(8.71/17.95 =)48.5\%$ of the summed redundancies of the regions at this geographical scale (NUTS-3).

One of the advantages of entropy statistics is that the values can be fully decomposed. Analogously to the decomposition of probabilistic entropy [16.120, 20f.], mutual redundancy in three (or more) dimensions can be decomposed into groups as follows

$$R = R_0 + \sum_G \frac{n_G}{N} R_G\,. \tag{16.9}$$

When one decomposes in the geographical dimension, R_0 represents redundancy generated between regions; R_G is the synergy generated at a geographical scale G; n_G is the number of firms at this geographical scale; and N the total number of firms in the aggregate ($N = 1\,187\,421$ in the Swedish case).

The between-group redundancy (R_0) can be considered as a measure of the synergy among regions. A negative value of R_0 indicates an additional synergy (i. e., redundancy generation) at the next level of national agglomeration among the lower-level geographical units. In the Netherlands and Sweden, for example, such a surplus was found at the national level; in Germany, this surplus was mainly found at the level of the federal states (*Länder*). Whereas one cannot compare the quantitative values of R_0 across countries—because these values are sample-specific—one is allowed to compare the indicator in terms of the positive or negative signs of R_0 and as percentages of the total synergy (R_{123}). The percentages of synergy which is generated above the regional level in the various country studies, for example, can be found in column (e) of Appendix 16.A.

Table 16.2 shows that in the case of Sweden, the surplus of the national system is -4.61 mbit (on top of the aggregation of the results at individual counties). This is 25.7% of the -22.56 mbit measured for Sweden as a national system. In other words, one quarter of the reduction of uncertainty in the national system is realized at a level higher than within the regions. At the next level of aggregation (NUTS2), an additional synergy of $(22.56-19.84) = 2.72$ mbit, or 13.7%, is realized. Among the three *Landsdelar* (NUTS1), however, only 0.5 mbit, or 2.2% of the national sum total, is reduced by this further aggregation. In other words, the Swedish national system is organized hierarchically, as indeed is suggested by most of the literature about Sweden.

Analogous to the geographical decomposition, one can also decompose redundancy in terms of industrial sectors or firm sizes. In a series of studies, we decomposed a number of national systems of innovation: Germany [16.122], the Netherlands [16.123], Sweden [16.121], Norway [16.124], Italy [16.125], Hungary [16.126], the Russian Federation [16.127], and China [16.128]. In the case of the Netherlands, Norway, Sweden, and China, the national level adds to the sum of the regions. In the Netherlands, the (inter-regional) highways to Amsterdam Airport (Schiphol) are probably the most important axes of the knowledge-based economy. In Sweden, the synergy is concentrated in three regions (Stockholm, Gothenburg, and Malmö/Lund); in China, four municipalities which are administered at the national level participate in the knowledge-based economy more than comparable regions. As noted, summary information about the various country studies is provided in the Appendices 16.A and 16.B. For more detailed information about data and methodology the reader is referred to the original studies. Here, we focus on the theoretical conclusions of these studies.

In Norway, foreign-driven investments along the west coast in the marine and maritime industries drive the transition from a political to a knowledge-based economy. The synergy in terms of the development of new options is larger in these coastal regions than in the regions with the traditional universities in Oslo and Trondheim. Hungary's western part is transformed by integration into the European Union, whereas the eastern part has remained a state-led innovation system. The capital Budapest occupies a separate position as a metropolitan system of innovations. The national level no longer adds synergy to the sum of the synergies in these three regional systems.

In a study of Italy, we first used the administrative units (NUTS2 regions) provided by Eurostat and the OECD. The data is the complete set of 4.5 million Italian firms registered at Statistics Italy in 2007. Figure 16.9 shows that the main division in this country is between the northern and southern parts of the country. (Sicily has a special position.) In other words, the pattern is opposite to the one for Sweden summarized in Table 16.2 above: the regions are administrative artifacts, while the country is organized in terms of two main innovation systems, each with a different dynamics. The aggregation among the lower level regions indicates the role of the national system, but this role is different in the northern and southern parts of Italy. The perspective on Italy in terms of regions [16.129] is not supported by these results.

One of the conclusions to be drawn throughout the studies of the more advanced economies, is that knowledge-intensive services do not contribute to the local synergy in regions because they are not coupled geographically. For example, if one offers a knowledge-intensive service in Munich and receives a phone call from Hamburg, the next step is simply to take a plane to Hamburg, or perhaps to catch a high-speed train. In other words, it does not matter whether one is located in Munich or Hamburg as knowledge-intensive services uncouple from the local economy. The main competitive advantage is proximity to an airport or train station.

In a study of the Russian Federation, however, the national level could be shown to disorganize synergy development at lower levels. Knowledge-intensive services (KIS) cannot sufficiently circulate in Russia because of their integration into the (localized) state apparatuses. Relative *foot-looseness* [16.130] of KIS can also be expected in the case of high-tech knowledge-

Table 16.2 Between-group synergy at different geographical scales in the Swedish innovation system. Source: [16.121]

Geographical scale	$\sum R$ (mbit)	R_0	R_0 as % contribution
NUTS0 (national level)	−22.56		
NUTS1 (3 Landsdelar)	−22.08	−0.48	2.2
NUTS2 (8 Riksområden)	−19.84	−2.72	13.7
NUTS3 (21 Counties)	−17.95	−4.61	25.7

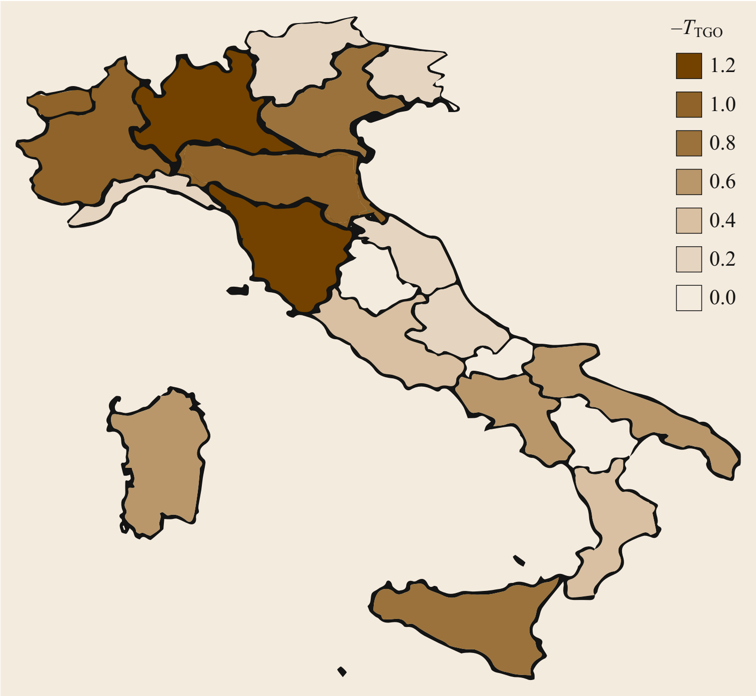


Fig. 16.9 Decomposition of Italy in terms of regions; n of firms $= 4\,480\,473$. Data: Statistics Italy (IStat), 2007. Source: [16.125]

based manufacturing; but the expectation is different for medium-tech manufacturing, because in this case the dynamics are often more embedded in other parts of the economy [16.131].

A number of policy implications follow from these conclusions and considerations. Footloose companies cannot be expected to contribute to the strengthening of the integration within a given region. High-tech knowledge-intensive services, however, may require a laboratory. One would expect medium-tech manufacturing to be more embedded and thus to generate more employment than high-tech.

In summary, the different country studies show that the patterns can be very different among nations as well as among regions within nations [16.132]. The dynamics are also different when comparing the sciences with markets: in publication systems uncoupling and international (that is, nonlocalized) orientation can be considered as improvements to the system, while in the case of regional developments the focus is on retaining *wealth from knowledge* and thus on developing local synergies. The discussion of the potential uncoupling from geographical locations by knowledge-intensive services illustrates how the different dynamics can also be interwoven. High-tech and knowledge-intensivity tend to induce globalization, including volatility, since stabilization is not a priority. The trade-off between the knowledge-based economy self-organizing at the global level and the lower-level organization in networked instantiations can be measured in considerable detail using the TH indicator.

Since the dynamics are complex, the results can be counter-intuitive because the a priori categories are being tested as hypotheses. Where are empirically the windows of opportunity for coupling self-organizing and differentiation with integration into organized forms and along trajectories? In a recent study of Spain, for example, Andalucia as a region (at the NUTS2 level) did poorly in generating mutual redundancy, but Sevilla as a town within this region (NUTS3) showed a different pattern [16.133]. In summary, one of the major functions of these studies is to revise and inform the categories used for making such assessments. Revision may make these categories more knowledge-based and thus enhance the visibility of new options.

16.8 Institutional Retention

Note that the TH indicator is a systemic indicator. Activities in a specific region (e. g., Linköping in Sweden; [16.134]) may have been very successful in terms of developing university–industry–government relations, but entrepreneurship is a form of action. One expects national governments and European policies to develop action plans to stimulate *less favored regions*. However, the TH-indicator informs us about the environments of these entrepreneurial activities. The chances of being successful as an innovative entrepreneur are statistically higher in Stockholm than in Linköping because of the relative reduction of uncertainty in the former region. This conclusion is not meant to discourage entrepreneurship in lagging regions. On the contrary, one may also conclude as a policy implication that some regions do not need support because the dynamics of the knowledge-based economy is already self-organizing in these regions.

Action, entrepreneurship, and local organization combine and integrate technical opportunities, market perspectives, and geographical resources (e. g., endowments). Selection mechanisms, however, differentiate on the basis of different criteria. Insofar as the criteria interact, redundancy may be generated. In other words, the complex dynamics is both differentiating and integrating. The neoevolutionary model focuses first on the *structural differentiation*: how is the system driven to change? The (neo-)institutional perspective on university–industry–government relations focuses on *integration in action*; for example, in terms of academic entrepreneurship.

From this perspective, the focus is on finding new ways to enhance innovation, such as the invention of venture capital. The cognitive dimension is endogenized into this model in the context of policy innovations. For example, priority programs at strategic research sites such as the emergence of new interdisciplinary fields (e. g., computer science) have been synthesized since the mid-20th century [16.135]. New fields are actively coconstructed as opposed to a previous model of branching into specialization [16.136].

Despite the intrinsic relations between the institutional and evolutionary models, the resulting research programs are different in important respects. A range of metrics have been developed from the institutional perspective on TH relations. These approaches do not present a single model for capturing *the Triple Helix*, but focus on different aspects of TH relations. Contributions link diverse themes ranging from conceptual work on entrepreneurial science (e. g., [16.2, 137]) and academic capitalism [16.138], or entrepreneurial universities [16.4] that can act as regional innovation organizers [16.139], to research on indicators such as university patenting and licensing [16.140, 141] and academic inventors [16.142, 143]. The theoretical frameworks of the empirical studies span a large domain including organization studies, business and management, network science, etc. It is beyond the scope of this chapter to review all these approaches which touch upon the TH theme.

We illustrate the diversity of theoretical backgrounds showing a coword map (Fig. 16.10) based on 139 keywords attributed five or more times to 492 documents retrieved from the Social Science Citation Index and the Arts & Humanities Citation Index on November 16, 2016 using the search string: `ts="Triple Helix" OR ts="university-industry-government relations"`. Seven clusters are distinguished, among them three major ones in red, green, and blue. `Triple-helix` as an adjective (with hyphen) is positioned in the blue-colored cluster on the left side labeled *entrepreneurial university*. `Triple Helix` as a substantive (without hyphen) is positioned in the green cluster on the right side labeled *systems dynamics*. The main cluster in red at the top relates the TH theme to studies of higher education and university policies. Scientometric indicators and social networks are indicated at the bottom with keywords that refer to new technologies such as *nanotechnology*.

A considerable number of contributions are concerned with capturing academic entrepreneurship [16.144]. Academic patenting has been debated in relation to the Bayh–Dole Act of 1980, which changed the intellectual property rights on academic inventions in the USA. *Mowery* et al. [16.145], for example, argued that the Bayh–Dole Act has been an important driver of university patenting and licensing activity. The entrepreneurial university has been another starting point for the development of indicators to measure Triple Helix relations [16.142, 146, 147]. *Narin* et al. [16.148, p. 317] considered the rapidly growing citation linkages between US patents and scientific literature as "useful evidence in arguing the case for governmental support of science."

Most of the scientometric contributions are method-driven; the TH is used as a metaphor in the theoretical background. Some authors argue for extending the metaphor to four or more helices, including for example the public, or the relation between developed and developing countries [16.149, 150]. The extension of the TH indicator of synergy to more than three dimensions is straightforward [16.119].

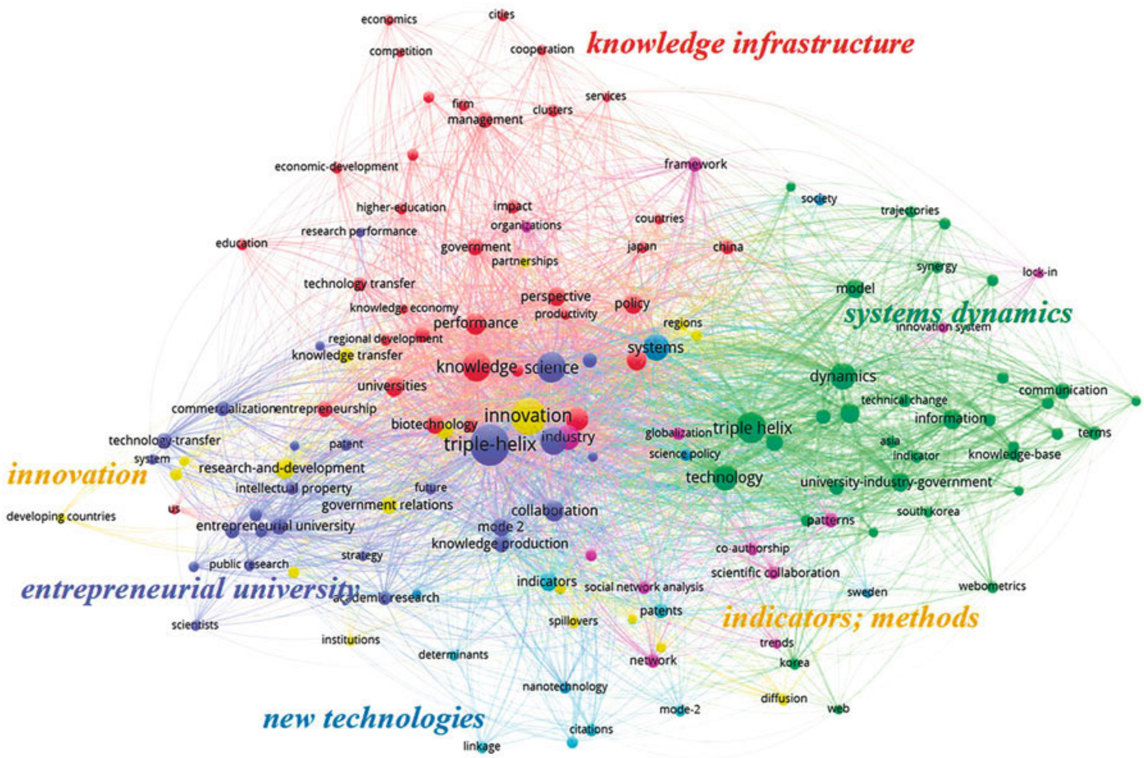


Fig. 16.10 Semantic map based on 139 keywords attributed five or more times to 492 documents retrieved from the Social Science Citation Index and the Arts & Humanities Citation Index on November 16, 2016; search string: `ts="Triple Helix" OR ts="university-industry-government relations"`. VOSviewer used for the layout and clustering

16.9 Concluding Remarks

The Triple Helix provides a metaphor which can be used in modeling the knowledge-based economy and innovation. The dynamics of a knowledge-based economy are complex [16.78, 151, 152]. We have argued that the challenges of modeling are not only theoretical. Systemness can be operationalized and measured in terms of the generation of new options. Without such operationalization, the notion of *system* tends to lead to reification; knowledge-based systems are not given, but constructed.

However, one can assume and test the possibility of emerging systemness. Can synergy be indicated and if so, at which level? From a communication-theoretical perspective, the not yet realized options can be considered as redundancy. Redundancy is developed by providing meanings from different perspectives to the same or similar events. A Triple Helix of university–industry–government relations provides these different perspectives.

In addition to horizontal differentiation among the three helices, we have operationalized vertical differentiation. The vertical differentiation finds its origin in the focus on relations in the neoinstitutional model [16.25, 27]: the nodes (in this case, the institutions) operate by relating; the relations relate in a second-order dynamics of possible relations. Meaning is provided from the perspective of hindsight to events invoking horizons of meaning that can be codified differently. The codes are generated by what has also been called institutional logics [16.153] or they can be considered as the eigenvectors in a vector-space [16.154]. A three-layered system is thus envisaged: information processing in relations at the bottom; meaning processing in a vector space based on correlations among distributions of relations; and thirdly, an interaction between meaning processing and codes of communication that opens horizons in which cognitive distinctions can be constructed. In other words, these cognitive constructs are embedded in social constructions such as networks and institutions.

New options are developed through cultural practices. The social structures carry the knowledge-based structures which are themselves carried by reflexive agents who infrareflexively have access to and can integrate representations in all three layers [16.81]. Agents are thus able to change structures; the structures mediate in layers of communication which transform events (including actions) into expectations and expectations into discursive knowledge [16.155]. Construction is bottom-up; control tends to be top-down. As a new selection environment is constructed, the locus of control can be expected to shift. As yet another selection environment, the knowledge base thus transforms the political economy in which it remains under construction.

Acknowledgments. We wish to thank Henry Etzkowitz, Han Woo Park, two anonymous referees, and the editors of the SPRU Working Papers Series (SWPS) for comments on previous versions.

16.A Appendix: Comparison Among Country Studies in Terms of the Main Results

Table 16.3 Appendix I: Comparison among country studies in terms of the main results

Country (a)	*N* of firms (b)	Year (c)	Source (d)	R_0/R (%) (e)	R_0 between regions (f)	sum ΔR regions (g)	*R* national level[a] (h)
China							
31 provinces	379 026	2008–2010	ORBIS	18.0	−35.46	−161.02	−196.48
339 2nd level				77.7	−142.58	−40.84	−183.42
Germany	2 119 028	2002	Social Insurance Statistics	10.1	−18.17	−161.91	−180.08
Hungary	659 701	Dec-05	Statistics Hungary	−46.5	*10.94*	−34.48	−23.55
Italy	4 480 473	2007	Statistics Italy	63.1	−15.63	−9.15	−24.78
	462 316	2011	Amadeus	22.9	−15.14	−50.96	−66.10
Netherlands	1 131 668	2001	Chambers of Commerce	21.3	−9.09	−33.55	−42.64
Norway	481 819	2008	Statistics Norway	11.7	−11.68	−87.92	−99.50
Russian Federation	593 987	2011	ORBIS	37.9	−1019.8	−1670.9	−2690.7
Spain	1 011 016	2010	ORBIS	12.0	−106.00	−780	−886.00
Sweden	1 187 421	Nov. 2011	Statistics Sweden	20.4	−4.61	−17.95	−22.56
USA [16.156]							
States	8 121 301	2016	ORBIS	2.8	−2.20	−76.70	−78.88
CBSA				38.6	−30.39	−48.35	−78.74

[a] *R* is redundancy

16.B Appendix: Comparison Among Country Studies in Terms of the Data

Table 16.4 Appendix II: Comparison among country studies in terms of the data

Country	Regions (i)	Recall (j)
China	31 provinces 339 admin. units at the 2nd level	47.6% of the retrieval (idem)
Germany	Data is aggregated at the level of 438 districts (NUTS-3)	Entire population
Hungary	19 NUTS-3 units (counties)	54% of the retrieval
Italy	20 regions at the NUTS-2 level 20 regions at the NUTS-2 level	Entire population 46.6% of the retrieval
Netherlands	40 COROP regions at the NUTS-3 level	Entire population
Norway	19 counties at the NUTS-3 level	Entire population
Russian Federation	83 Federal Subjects or States	96.9% of the retrieval
Spain	51 provinces at the NUTS-3 level.	40.1% of the retrieval
Sweden	21 counties at the NUTS-3 level; Swedish Sector specification	Entire population
USA States [16.156] CBSA	50+ States 945+ CBSA regions	95.6% of the retrieval (idem)

References

16.1 H. Etzkowitz: Entrepreneurial scientists and entrepreneurial universities in American academic science, Minerva **21**(2/3), 198–233 (1983)

16.2 H. Etzkowitz: The norms of entrepreneurial science: Cognitive effects of the new university-industry linkages, Res. Policy **27**(8), 823–833 (1998), https://doi.org/10.1016/S0048-7333(98)00093-6

16.3 H. Etzkowitz: *MIT and the Rise of Entrepreneurial Science* (Routledge, London 2002)

16.4 B.R. Clark: *Creating Entrepreneurial Universities: Organization Pathways of Transformation* (Pergamon, Guildford 1998)

16.5 L. Leydesdorff, P. Van den Besselaar (Eds.): *Evolutionary Economics and Chaos Theory: New Directions in Technology Studies* (Pinter, London, New York 1994)

16.6 G. Dosi: Technological paradigms and technological trajectories: A suggested interpretation of the determinants and directions of technical change, Res. Policy **11**(3), 147–162 (1982)

16.7 G. Bianconi, R.K. Darst, J. Iacovacci, S. Fortunato: Triadic closure as a basic generating mechanism of communities in complex networks, Phys. Rev. E **90**(4), 042806 (2014)

16.8 G. Simmel: Quantitative aspects of the group. In: *The Sociology of Georg Simmel*, ed. by K.H. Wolff (Free Press, Glencoe 1950) pp. 87–177

16.9 H. Etzkowitz, L. Leydesdorff: The dynamics of innovation: From national systems and 'Mode 2' to a triple helix of university-industry-government relations, Res. Policy **29**(2), 109–123 (2000)

16.10 I.A. Ivanova, L. Leydesdorff: Redundancy generation in university-industry-government relations: The triple helix modeled, measured, and simulated, Scientometrics **99**(3), 927–948 (2014)

16.11 I.A. Ivanova, L. Leydesdorff: Rotational symmetry and the transformation of innovation systems in a triple helix of university-industry-government relations, Technol. Forecast. Soc. Change **86**, 143–156 (2014)

16.12 H.A. Simon: The organization of complex systems. In: *Hierarchy Theory: The Challenge of Complex Systems*, ed. by H.H. Pattee (George Braziller, New York 1973) pp. 1–27

16.13 G.S. Brown: *Laws of Form* (George Allen and Unwin, London 1969)

16.14 A. Petersen, D. Rotolo, L. Leydesdorff: A triple helix model of medical innovations: Supply, demand, and technological capabilities in terms of medical subject headings, Res. Policy **45**(3), 666–681 (2016)

16.15 R.R. Nelson, K. Buterbaugh, M. Perl, A. Gelijns: How medical know-how progresses, Res. Policy **40**(10), 1339–1344 (2011)

16.16 M. Saad, G. Zawdie: *Theory and Practice of Triple Helix Model in Developing Countries: Issues and Challenges* (Routledge, New York 2011)

16.17 Y. Li, S. Arora, J. Youtie, P. Shapira: Using web mining to explore triple helix influences on growth in small and mid-size firms, Technovation (2016), https://doi.org/10.1016/j.technovation.2016.01.002

16.18 H. Etzkowitz: *The Triple Helix: University-Industry-Government Innovation in Action* (Routledge, London 2008)

16.19 H. Etzkowitz, C. Zhou: *Triple Helix: University-Industry-Government Innovation and Entrepreneurship* (Routledge, Milton Park 2017)

16.20 C. Zhou, X.-M. Peng: The entrepreneurial university in China: Nonlinear paths, Sci. Public Policy **35**(9), 637–646 (2008)

16.21 H. Etzkowitz, L. Leydesdorff (Eds.): *Universities in the Global Knowledge Economy* (Cassell Academic, London 1997)

16.22 H. Etzkowitz, L. Leydesdorff: The endless transition: A "Triple Helix" of university-industry-government relations, introduction to a theme issue, Minerva **36**, 203–208 (1998)
16.23 L. Leydesdorff, I.A. Ivanova: Mutual redundancies in inter-human communication systems: Steps towards a calculus of processing meaning, J. Assoc. Inf. Sci. Technol. **65**(2), 386–399 (2014)
16.24 C.G. Langton: Artificial Life. In: *Artificial Life*, Vol. VI, ed. by C.G. Langton (Addison-Wesley, Redwood 1989) pp. 1–47
16.25 J.F. Padgett, W.W. Powell: The problem of emergence. In: *The Emergence of Organizations and Markets*, ed. by J.F. Padgett, W.W. Powell (Princeton Univ. Press, New Jersey 2012) pp. 1–32
16.26 M. Storper: *The Regional World—Territorial Development in a Global Economy* (Guilford, New York 1997)
16.27 W.W. Powell, P.J. DiMaggio (Eds.): *The New Institutionalism in Organizational Analysis* (Univ. of Chicago Press, Chicago 1991)
16.28 L. Leydesdorff: The knowledge-based economy and the triple helix model, Annu. Rev. Inf. Sci. Technol. **44**, 367–417 (2010)
16.29 J. Schumpeter: *Business Cycles: A Theoretical, Historical and Statistical Analysis of Capitalist Process* (McGraw-Hill, New York 1964)
16.30 L. Leydesdorff: Epilogue. In: *Evolutionary Economics and Chaos Theory: New Directions for Technology Studies*, ed. by L. Leydesdorff, P. v.d. Besselaar (Pinter, London, New York 1994) pp. 180–192
16.31 L. Leydesdorff, P. van den Besselaar: Technological development and factor substitution in a non-linear model, J. Soc. Evolut. Syst. **21**, 173–192 (1998)
16.32 J.A. Schumpeter: *The Theory of Economic Development* (Harvard Univ. Press, Cambridge 1949)
16.33 L. Soete, B. ter Weel: *Schumpeter and the knowledge-based economy: On technology and competition policy*, Research Memoranda, Vol. 004 (MERIT (Maastricht Economic Research Institute on Innovation and Technology), Maastricht 1999)
16.34 P. Cooke, L. Leydesdorff: Regional development in the knowledge-based economy: The construction of advantages, J. Technol. Transf. **31**(1), 5–15 (2006)
16.35 C. Freeman, C. Perez: Structural crises of adjustment, business cycles and investment behaviour. In: *Technical Change and Economic Theory*, ed. by G. Dosi, C. Freeman, R. Nelson, G. Silverberg, L. Soete (Pinter, London 1988) pp. 38–66
16.36 W.W. Powell: Neither market nor hierarchy: Network forms of organization, Res. Organ. Behav. **12**, 295–336 (1990)
16.37 R.R. Nelson: Recent evolutionary theorizing about economic change, J. Econ. Lit. **33**(1), 48–90 (1995)
16.38 R.R. Nelson, S.G. Winter: *An Evolutionary Theory of Economic Change* (Belknap Press of Harvard Univ. Press, Cambridge 1982)
16.39 H.-J. Braczyk, P. Cooke, M. Heidenreich (Eds.): *Regional Innovation Systems* (Univ. College London Press, London, Bristol 1998)
16.40 B. Carlsson: Internationalization of innovation systems: A survey of the literature, Res. Policy **35**(1), 56–67 (2006)
16.41 A. Giddens: *Central Problems in Social Theory* (Macmillan, London 1979)
16.42 K. Marx: *Grundrisse: Foundations of the Critique of Political Economy (Rough Draft)* (Penguin, Harmondsworth 1973)
16.43 K. Marx: *Das Kapital I* (Dietz, Berlin 1971)
16.44 J.J. Beer: *The Emergence of the German Dye Industry* (University of Illinois Press, Urbana 1959)
16.45 H. Braverman: *Labor and Monopoly Capital. The Degradation of Work in the Twentieth Century* (Monthly Review, New York, London 1974)
16.46 D. Noble: *America by Design* (Knopf, New York 1977)
16.47 M. Aoki: *Towards a Comparative Institutional Analysis* (MIT Press, Cambridge 2001)
16.48 P.M. Allen: Evolutionary complex systems: Models of technology change. In: *Evolutionary Economics and Chaos Theory: New Directions in Technology Studies*, ed. by L. Leydesdorff, P. van den Besselaar (Pinter, London 1994) pp. 1–18
16.49 P. Dasgupta, P.A. David: Toward a new economics of science, Res. Policy **23**(5), 487–521 (1994)
16.50 F. Malerba, R. Nelson, L. Orsenigo, S. Winter: 'History-friendly' models of industry evolution: The computer industry, Ind. Corp. Change **8**(1), 3–35 (1999)
16.51 G. Hodgson, T. Knudsen: *Darwin's Conjecture: The Search for General Principles of Social and Economic Evolution* (Univ. of Chicago Press, Chicago, London 2011)
16.52 E.S. Andersen: *Evolutionary Economics: Post-Schumpeterian Contributions* (Pinter, London 1994)
16.53 B.R. Martin: The evolution of science policy and innovation studies, Res. Policy **41**(7), 1219–1239 (2012)
16.54 F. List: *The National Systems of Political Economy* (Longman, London 1904)
16.55 C. Freeman: *Technology, Policy, and Economic Performance: Lessons from Japan* (Pinter, London 1987)
16.56 B.-Å. Lundvall: Innovation as an interactive process: From user-producer interaction to the national system of innovation. In: *Technical Change and Economic Theory*, ed. by G. Dosi, C. Freeman, R. Nelson, G. Silverberg, L. Soete (Pinter, London 1988) pp. 349–369
16.57 R.R. Nelson (Ed.): *Government and Technical Progress: A Cross-Industry Analysis* (Pergamon, New York 1982)
16.58 P.A. Hall, D.W. Soskice (Eds.): *Varieties of Capitalism: The Institutional Foundations of Comparative Advantage* (Oxford Univ. Press, Oxford 2001)
16.59 R.R. Nelson, S.G. Winter: In search of useful theory of innovation, Res. Policy **6**, 35–76 (1977)

16.60 N. Rosenberg: The direction of technological change: Inducement mechanisms and focusing devices. In: *Perspectives on Technology* (Cambridge Univ. Press, Cambridge 1976) pp. 108–125
16.61 N. Rosenberg: *Inside the Black Box: Technology and Economics* (Cambridge Univ. Press, Cambridge 1982)
16.62 T.J. Pinch, W.E. Bijker: The social construction of facts and artifacts: Or how the sociology of science and the sociology of technology might benefit each other, Soc. Stud. Sci. **14**, 399–442 (1984)
16.63 F.W. Geels: Technological transitions as evolutionary reconfiguration processes: A multi-level perspective and a case-study, Res. Policy **31**(8/9), 1257–1274 (2002)
16.64 E.S. Andersen: *Artificial Economic Evolution and Schumpeter* (Institute for Production, University of Aalborg, Aalborg 1992)
16.65 Z. Griliches: Productivity, R&D and the data constraint, Am. Econ. Rev. **84**(1), 1–23 (1994)
16.66 Z. Griliches: Introduction. In: *Output measurement in the service sectors*, National Bureau of Economic Research Studies in Income and Wealth, Vol. 56, ed. by Z. Griliches (University of Chicago Press, Chicago 1992) pp. 1–22
16.67 K. Pavitt: Sectoral patterns of technical change: Towards a theory and a taxonomy, Res. Policy **13**(6), 343–373 (1984)
16.68 OECD/Eurostat: *Proposed Guidelines for Collecting and Interpreting Innovation Data, "Oslo Manual"* (OECD, Paris 1997)
16.69 A.B. Jaffe, M. Trajtenberg: *Patents, Citations, and Innovations: A Window on the Knowledge Economy* (MIT Press, Cambridge, London 2002)
16.70 D. Sahal: Alternative conceptions of technology, Res. Policy **10**(1), 2–24 (1981)
16.71 D. Foray, B.-A. Lundvall: The knowledge-based economy: From the economics of knowledge to the learning economy. In: *Employment and Growth in the Knowledge-Based Economy*, ed. by D. Foray, B.-A. Lundvall (OECD, Paris 1996) pp. 11–32
16.72 A.P. Carter: Measuring the performance of a knowledge-based economy. In: *Employment and Growth in the Knowledge-Based Economy*, ed. by D. Foray, B.-Å. Lundvall (OECD, Paris 1996) pp. 61–68
16.73 B. Godin: The knowledge-based economy: Conceptual framework or buzzword?, J. Technol. Transf. **31**(1), 17–30 (2006)
16.74 P.A. David, D. Foray: Assessing and expanding the science and technology knowledge base, STI Rev. **16**, 13–68 (1995)
16.75 P.A. David, D. Foray: An introduction to the economy of the knowledge society, Int. Soc. Sci. J. **54**(171), 9–23 (2002)
16.76 R. Cowan, D. Foray: The economics of codification and the diffusion of knowledge, Industrial and Corporate Change **6**(3), 595–622 (1997)
16.77 R. Cowan, P. David, D. Foray: The explicit economics of knowledge codification and tacitness, Industrial and Corporate Change **9**(2), 211–253 (2000)
16.78 N. Luhmann: *Systemtheorie, Evolutionstheorie und Kommunikationstheorie*, Soziologische Aufklärung, Vol. 2 (Westdeutscher Verlag, Opladen 1975) pp. 193–203, Reprinted in: *The Differentiation of Society, Translated by Stephen Holmes and Charles Larmore* (Columbia University Press, New York 1982) pp. 255–270
16.79 D. Sahal: Technological guideposts and innovation avenues, Res. Policy **14**, 61–82 (1985)
16.80 M. Ashmore: *The Reflexivity Thesis: Wrighting Sociology of Scientific Knowledge* (University of Chicago Press, Chicago, London 1989)
16.81 B. Latour: The politics of explanation: An alternative. In: *Knowledge and Reflexivity: New Frontiers in the Sociology of Knowledge*, ed. by S. Woolgar, M. Ashmore (SAGE, London 1988) pp. 155–177
16.82 P. Wouters: *The Citation Culture*, Ph.D. Thesis (Univ. of Amsterdam, Amsterdam 1999)
16.83 W. Krohn, E.T. Layton, P. Weingart (Eds.): *The Dynamics of Science and Technology: Social Values, Technical Norms, and Scientific Criteria in the Development of Knowledge*, Vol. 2 (Reidel, Dordrecht 1978)
16.84 W. van den Daele, P. Weingart: Resistenz und Rezeptivität der Wissenschaft – zu den Entstehungsbedingungen neuer Disziplinen durch wissenschaftliche und politische Steuerung, Z. Soziol. **4**(2), 146–164 (1975)
16.85 E. Husserl: *Cartesianische Meditationen und Pariser Vorträge* (Martinus Nijhoff, The Hague 1929), [Cartesian Meditations and the Paris Lectures, translated by Dorion Cairns 1960]
16.86 L. Leydesdorff: The sciences are discursive constructs: The communication perspective as an empirical philosophy of science. In: *Communication and Technology*, ed. by L. Cantoni, J.A. Danowski (De Gruyter Mouton, Berlin, Boston 2015) pp. 553–562
16.87 W. Glänzel, A. Schubert: Analysing scientific networks through co-authorship. In: *Handbook of Quantitative Science and Technology Research*, ed. by H.F. Moed, W. Glänzel, U. Schmoch (Springer, Dordrecht 2005) pp. 257–276
16.88 O. Persson, W. Glänzel, R. Danell: Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies, Scientometrics **60**(3), 421–432 (2004)
16.89 C.S. Wagner, L. Leydesdorff: Network structure, self-organization and the growth of international collaboration in science, Res. Policy **34**(10), 1608–1618 (2005)
16.90 W. Van den Daele, W. Krohn, P. Weingart (Eds.): *Geplante Forschung: Vergleichende Studien über den Einfluss politischer Programme auf die Wissenschaftsentwicklung* (Suhrkamp, Frankfurt a.M. 1979)
16.91 M. Callon, J. Law, A. Rip (Eds.): *Mapping the Dynamics of Science and Technology* (Macmillan, London 1986)

16.92 H. Small: Cited documents as concept symbols, Soc. Stud. Sci. **8**(3), 113–122 (1978)
16.93 P. Wouters: The signs of science, Scientometrics **41**(1), 225–241 (1998)
16.94 L. Leydesdorff: *The Challenge of Scientometrics: The Development, Measurement, and Self-Organization of Scientific Communications* (DSWO Press, Leiden Univ., Leiden 1995)
16.95 R.R. Braam, H.F. Moed, A.F.J. van Raan: Mapping of science by combined co-citation and word analysis. I. Structural aspects, J. Assoc. Inf. Sci. Technol. **42**(4), 233–251 (1991)
16.96 R.R. Braam, H.F. Moed, A.F.J. van Raan: Mapping of science by combined co-citation and word analysis. II: Dynamical aspects, J. Assoc. Inf. Sci. Technol. **42**(4), 252–266 (1991)
16.97 Y. Fujigaki, L. Leydesdorff: Quality control and validation boundaries in a triple helix of university-industry-government: "Mode 2" and the future of university research, Soc. Sci. Inf. **39**(4), 635–655 (2000)
16.98 E. Bruckner, W. Ebeling, M.J. Montaño, A. Scharnhorst: Nonlinear stochastic effects of substitution – An evolutionary approach, J. Evol. Econ. **6**(1), 1–30 (1996)
16.99 P.O. Seglen: The skewness of science, J. Am. Soc. Inf. Sci. **43**(9), 628–638 (1992)
16.100 T.W. Deacon: *Incomplete Nature: How the Mind Emerged from Matter* (Norton & Company, New York, London 2012)
16.101 C.E. Shannon: A Mathematical Theory of Cryptography, Memorandum MM-45-110-92 (AT&T Bell Laboratories, Murray Hill 1945)
16.102 C. Freeman, L. Soete: *The Economics of Industrial Innovation*, 3rd edn. (Pinter, London 1997)
16.103 K. Krippendorff: Information of interactions in complex systems, Int. J. Gen. Syst. **38**(6), 669–680 (2009)
16.104 K. Krippendorff: W. Ross Ashby's information theory: A bit of history, some solutions to problems, and what we face today, Int. J. Gen. Syst. **38**(2), 189–212 (2009)
16.105 R.E. Ulanowicz: The dual nature of ecosystem dynamics, Ecol. Model. **220**(16), 1886–1892 (2009)
16.106 S.A. Kauffman: *Investigations* (Oxford Univ. Press, Oxford 2000)
16.107 R.E. Rice, S.D. Cooper: *Organizations and Unusual Routines: A Systems Analysis of Dysfunctional Feedback Processes* (Cambridge Univ. Press, Cambridge 2010)
16.108 K. Frenken, L. Leydesdorff: Scaling trajectories in civil aircraft (1913–1970), Res. Policy **29**(3), 331–348 (2000)
16.109 L. Leydesdorff: The mutual information of university-industry-government relations: An indicator of the triple helix dynamics, Scientometrics **58**(2), 445–467 (2003)
16.110 H.W. Park, L. Leydesdorff: Longitudinal trends in networks of university-industry-government relations in South Korea: The role of programmatic incentives, Res. Policy **39**(5), 640–649 (2010)
16.111 R.E. Ulanowicz: *Growth and Development: Ecosystems Phenomenology* (toExcel, San Jose 1986)
16.112 N. Abramson: *Information Theory and Coding* (McGraw-Hill, New York 1963)
16.113 W.J. McGill: Multivariate information transmission, Psychometrika **19**(2), 97–116 (1954)
16.114 F.Y. Ye, S.S. Yu, L. Leydesdorff: The triple helix of university-industry-government relations at the country level, and its dynamic evolution under the pressures of globalization, J. Am. Soc. Inf. Sci. Technol. **64**(11), 2317–2325 (2013)
16.115 H.W. Park, H.D. Hong, L. Leydesdorff: A comparison of the knowledge-based innovation systems in the economies of South Korea and the Netherlands using triple helix indicators, Scientometrics **65**(1), 3–27 (2005)
16.116 Y. Zhang, K. Chen, G. Zhu, R.C. Yam, J. Guan: Interorganizational scientific collaborations and policy effects: An ego-network evolutionary perspective of the Chinese Academy of Sciences, Scientometrics **108**(3), 1383–1415 (2016)
16.117 S. Choi, J.S. Yang, H.W. Park: Quantifying the Triple Helix relationship in scientific research: Statistical analyses on the dividing pattern between developed and developing countries, Qual. Quant. **49**(4), 1381–1396 (2015)
16.118 S. Choi, J.S. Yang, H.W. Park: The triple helix and international collaboration in science, J. Assoc. Inf. Sci. Technol. **66**(1), 201–212 (2015)
16.119 L. Leydesdorff, H.W. Park, B. Lengyel: A routine for measuring synergy in university–industry–government relations: Mutual information as a triple-helix and quadruple-helix indicator, Scientometrics **99**(1), 27–35 (2014)
16.120 H. Theil: *Statistical Decomposition Analysis* (North-Holland, Amsterdam, London 1972)
16.121 L. Leydesdorff, Ø. Strand: The Swedish system of innovation: Regional synergies in a knowledge-based economy, J. Am. Soc. Inf. Sci. Technol. **64**(9), 1890–1902 (2013)
16.122 L. Leydesdorff, M. Fritsch: Measuring the knowledge base of regional innovation systems in germany in terms of a triple helix dynamics, Res. Policy **35**(10), 1538–1553 (2006)
16.123 L. Leydesdorff, W. Dolfsma, G. van der Panne: Measuring the knowledge base of an economy in terms of Triple-Helix relations among 'technology, organization, and territory', Research Policy **35**(2), 181–199 (1991)
16.124 Ø. Strand, L. Leydesdorff: Where is synergy in the Norwegian innovation system indicated? Triple helix relations among technology, organization, and geography, Technol. Forecast. Soc. Change **80**(3), 471–484 (2013)
16.125 I. Cucco, L. Leydesdorff: *Measuring Synergies in the Italian Innovation System: A Comparison of Amadeus™ and Administrative Data* (University Frederico, Naples 2013), unpublished manuscript
16.126 B. Lengyel, L. Leydesdorff: Regional innovation systems in Hungary: The failing synergy at the national level, Regional Studies **45**(5), 677–693 (2011)

16.127 L. Leydesdorff, E. Perevodchikov, A. Uvarov: Measuring triple-helix synergy in the Russian innovation systems at regional, provincial, and national levels, J. Assoc. Inf. Sci. Technol. **66**(6), 1229–1238 (2015)
16.128 L. Leydesdorff, P. Zhou: Measuring the Knowledge-Based Economy of China in terms of Synergy among Technological, Organizational, and Geographic Attributes of Firms, Scientometrics **98**(3), 1703–1719 (2014)
16.129 OECD: *OECD Reviews of Regional Innovation: Piedmont, Italy* (OECD, Paris 2009)
16.130 R. Vernon: The product cycle hypothesis in a new international environment, Oxf. Bull. Econ. Stat. **41**(4), 255–267 (1979)
16.131 W.M. Cohen, D.A. Levinthal: Absorptive capacity: A new perspective on learning and innovation, Adm. Sci. Q. **35**(1), 128–153 (1990)
16.132 J. Yoon, H.W. Park: Triple helix dynamics of South Korea's innovation system: A network analysis of inter-regional technological collaborations, Qual. Quant. **51**, 1–19 (2017)
16.133 L. Leydesdorff, I. Porto-Gomez: Measuring the expected synergy in spanish regional and national systems of innovation, J. Technol. Transf. **44**(1), 189–209 (2019)
16.134 H. Etzkowitz, M. Klofsten: The innovating region: Toward a theory of knowledge-based regional development, R&D Manag. **35**(3), 243–255 (2005)
16.135 H. Etzkowitz, L.S. Peters: Profiting from knowledge: Organisational innovations and the evolution of academic norms, Minerva **29**(2), 133–166 (1991)
16.136 J. Ben-David, R. Collins: Social factors in the origins of a new science: The case of psychology, Am. Sociol. Rev. **31**(4), 451–465 (1966)
16.137 H. Etzkowitz: Innovation in innovation: The triple helix of university-industry-government relations, Soc. Sci. Inf. **42**(3), 293–337 (2003)
16.138 S. Slaughter, L. Leslie: *Academic capitalism: Politics, Policies, and the Entrepreneurial University* (Johns Hopkins Univ. Press, Baltimore 1997)
16.139 H. Etzkowitz, A. Webster, C. Gebhardt, B. Terra: The future of the university and the university of the future: Evolution of ivory tower to entrepreneurial paradigm, Res. Policy **29**(2), 313–330 (2000)
16.140 OECD: *Compendium of Patent Statistics* (OECD, Paris 2005), https://www.timeshighereducation.com/news/oecd-compendium-of-patent-statistics-2005-link/201202.article
16.141 S. Saragossi, B. von Pottelsberghe: What patent data reveal about universities: The case of Belgium, J. Technol. Transf. **28**(1), 47–51 (2003)
16.142 M. Meyer, T. Sinilainen, J. Utecht: Towards hybrid triple helix indicators: A study of university-related patents and a survey of academic inventors, Scientometrics **58**(2), 321–350 (2003)
16.143 H. Etzkowitz: The entrepreneurial university: Vision and metrics, Ind. High. Educ. **30**(2), 83–97 (2016)
16.144 M. Meyer, K. Grant, P. Morlacchi, D. Weckowska: Triple helix indicators as an emergent area of enquiry: A bibliometric perspective, Scientometrics **99**(1), 151–174 (2014)
16.145 D.C. Mowery, R.R. Nelson, B.N. Sampat, A.A. Ziedonis: The growth of patenting and licensing by US universities: An assessment of the effects of the Bayh–Dole act of 1980, Res. Policy **30**(1), 99–119 (2001)
16.146 L. Liang, L. Chen, Y. Wu, J. Yuan: The role of chinese universities in enterprise-university research collaboration, Scientometrics **90**(1), 253–269 (2012)
16.147 L. Ranga, K. Debackere, N. von Tunzelmann: Entrepreneurial universities and the dynamics of academic knowledge production: A case study of basic vs. applied research in Belgium, Scientometrics **58**(2), 301–320 (2003)
16.148 F. Narin, K. Hamilton, D. Olivastro: The increasing linkage between US technology and public science, Res. Policy **26**(3), 317–330 (1997)
16.149 E.G. Carayannis, D.F.J. Campbell: Triple helix, quadruple helix and quintuple helix and how do knowledge, innovation, and environment relate to each other?, Int. J. Soc. Ecol. Sust. Develop. **1**(1), 41–69 (2010)
16.150 H.W. Park: Transition from the triple helix to n-tuple helices? An interview with Elias G. Carayannis and David F.J. Campbell, Scientometrics **99**(1), 203–207 (2014)
16.151 L. Leydesdorff, A. Petersen, I. Ivanova: The self-organization of meaning and the reflexive communication of information, Soc. Sci. Inf. **56**(1), 4–27 (2017)
16.152 H.A. Simon: The architecture of complexity, Proc. Am. Philos. Soc. **106**(6), 467–482 (1962)
16.153 P.H. Thornton, W. Ocasio, M. Lounsbury: *The Institutional Logics Perspective: A New Approach to Culture, Structure, and Process* (Oxford Univ. Press, Oxford 2012)
16.154 H. von Foerster: On self-organizing systems and their environments. In: *Self-Organizing Systems*, ed. by M.C. Yovits, S. Cameron (Pergamon, London 1960) pp. 31–50
16.155 L. Leydesdorff, M. Johnson, I.A. Ivanova: Mutual redundancies in inter-human communication systems: Steps towards a calculus of processing meaning, J. Assoc. Inf. Sci. Technol. **69**(10), 1181–1192 (2018)
16.156 L. Leydesdorff, C.S. Wagner, I. Porto-Gomez, J. A. Comins, F. Phillips: Synergy in the knowledge base of U.S. innovation systems at national, state, and regional levels: The contributions of high-tech manufacturing and knowledge-intensive services. J. Assoc. Inf. Sci. Technol. (2019), forthcoming

Loet Leydesdorff

Amsterdam School of Communication Research (ASCoR)
University of Amsterdam
Amsterdam, The Netherlands
loet@leydesdorff.net

Loet Leydesdorff (PhD Sociology, MA Philosophy, MSc Biochemistry) is Professor Emeritus at Amsterdam School of Communication Research (University of Amsterdam). He is Associate Faculty at the Science and Technology Policy Research Unit of the University of Sussex, Visiting Professor of the Institute of Scientific and Technical Information of China, and Guest Professor at Zhejiang University. He has published extensively in systems theory, social network analysis, scientometrics, and sociology of innovation. Clarivate Analytics lists him as highly-cited researcher since 2014.

Inga Ivanova

Institute for Statistical Studies and Economics of Knowledge
National Research University Higher School of Economics (NRU HSE)
Moscow, Russian Federation
inga.iva@mail.ru

Inga Ivanova (PhD Economics) is Research Fellow at the Institute for Statistical Studies and Economics of Knowledge's Centre for S&T, Innovation and Information Policy of the Higher School of Economics. She is Associate Professor at the Department of Educational Programmes of the Higher School of Economics and at the Far Eastern Federal University (Management Department). Research interests include innovation studies, network analysis, systems theory, and informational communication in social systems.

Martin Meyer

Kent Business School
University of Kent
Canterbury, UK
m.s.meyer@kent.ac.uk

Martin Meyer is Professor of Business & Innovation and Director of Kent Business School, University of Kent, Canterbury (UK). He has also held positions at the University of Sussex, Katholieke Universiteit Leuven, Helsinki University of Technology – now Aalto University, and Linköping University. He is well known for his work on science, technology and innovation indicators. His current research focuses on knowledge exchange, global innovation environments, and science-based technologies.

Science S Part C

Part C Science Systems and Research Policy

17. Scientometrics Shaping Science Policy and vice versa, the ECOOM Case

Koenraad Debackere, Wolfgang Glänzel, Bart Thijs

It is difficult to imagine a world without science policy. Ever since Vannevar Bush published his seminal insights on the role of science in society, science policy has become deeply ingrained in public policy. Alongside this, the discipline of scientometrics developed. It started from library and information needs, helping the ever-growing scientific community to access, retrieve and disseminate its ever-increasing output. However, along the way, scientometrics developed into a powerful set of scientifically validated data, indicators and tools. It diffused across many disciplines in the social sciences. Over time, this evolution came to the attention of policymakers. The wealth of data and indicators developed in the field of scientometrics (later extended to informetrics and webometrics) elicited interest in their use for policy purposes. A symbiosis between scientometrics and science policy was born. Using the case of the Flemish Centre for Research & Development Monitoring (ECOOM), we describe and illustrate this coevolution between scientometrics and science policy, its opportunities and its challenges, and its do's and don'ts.

17.1 Scientometrics and Science Policy, a Symbiotic Relationship

Since the seventeenth century, journals have been the most important communication channel in biomedical, life and natural sciences, as well as in a growing number of social sciences. They offer a reliable and valid source of data on scientific methods and results so long as they are peer reviewed, i. e., their content is endorsed by means of expert qualitative judgments. Critical review by peers of the scientific insights developed and posited in papers became a central mechanism to check and support scientific progress since that era. Both the publication literature (and its channeling into scientometric or, synonymously, bibliometric applications) and the peer-review mechanism became cornerstones in the fields of science, technology and innovation policy.

Since the first half of the twentieth century, the bibliographic structure of journal papers has become highly predictable and standardized compared to other publication venues, making them the ideal target for the (semi) automatic extraction of metadata preliminary to their statistical analyses. *Derek de Solla Price* formalized the paradigm of quantitative science studies in his book *Little Science, Big Science* [17.1, p. xvi]:

> My approach will be to deal statistically, in a not very mathematical fashion, with general problems of the shape and size of science and the ground rules governing growth and behaviour of science-in-the-large … I shall not discuss any part of the detail of scientific discoveries, their use and interrelations. I shall not even discuss specific scientists … The method to be used is similar to that of thermodynamics, in which is discussed the behaviour of a gas under various conditions of temperature and pressure.

Bibliometrics would be just a counting game if matters of size were its only concern—at some point, quantity had to yield to quality, or in de Solla Price's

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_17

words (quoted in *The Encyclopaedia of Library and Information Science*):

> Who dares to balance one paper of Einstein on relativity against even a hundred papers by John Doe, Ph.D., on the elastic constant of the various timbers (one to a paper) of the forest of lower Basutoland?

The missing link was filled in the 1960s. The citation network developed by Eugene Garfield's science citation index (SCI) provided a welcome answer since: (1) authors cite in the bibliographies of their papers mostly (although not only) other documents/authors that have supplied some form of prior relevant art, e. g., tools, methods, ideas, reviews of previous studies, etc., and (2) the number of citations accrued to a document/author provides a true estimate of its usefulness to others, hence it is an indicator of its cognitive impact. These ideas were pioneered by *Garfield* [17.2] in his seminal paper *Citation Indexes for science: a new dimension in documentation through association of ideas*.

It must be mentioned that before the *invention* of the SCI, quantitative and qualitative analyses of science were performed along routes such as the compilation and rankings of eminent scientists and the descriptive statistics of bibliographic data for historical or library management purposes (statistical bibliography). The advent of the SCI and the possibility to track quantity and quality patterns in the network of published scientific literature marked the point of no return in bibliometrics, legitimizing its entrance into the science policy arena since the 1970s. Some reflections helped this legitimation. Peer review was judged to be subject to many (personal, social) biases whereas bibliometric measures are objective, or at least they seem to be. Hence, rankings of authors and institutions based on publication outputs and citation scores can complement the peer-review process in the allocation of scientific merit (tenure, promotions, funding, etc.) while the mapping of bibliographic links between scientific documents can help understand the cognitive structure and the dynamics of scientific communication.

Today, it is difficult to imagine a world without science policy. Ever since *Vannevar Bush* published and pushed his seminal insights [17.3] on the role of science in society, science policy has become deeply ingrained in government functions and functioning. Alongside the advent of science policy, the discipline of scientometrics developed and evolved. It started from library and information needs, helping the ever-growing scientific community as described by *Derek de Solla Price* [17.1] to access, retrieve and disseminate its ever-increasing output. However, as the library and information needs developed, scientometrics developed into a powerful set of scientifically validated data, indicators, methods and tools. Scientometrics thus became the cornerstone of the field of library and information science [17.4]. In addition, scientometrics diffused across many disciplines in the social sciences, such as the sociology of science, history of science and the economics of science. This diffusion is briefly summarized in Fig. 17.1.

Over time, this evolution drew the attention of policymakers. The wealth of data and indicators developed in the field of scientometrics (later extended to informetrics and webometrics) elicited interest in their policy use and usability. A symbiosis between scientometrics and science policy was born, as summarized in Fig. 17.2. This symbiosis has led to multiple uses of scientometrics in the policy scene [17.5, 6]. As a consequence, scientometrics has been shaping science policy, while science policy is shaping bibliometrics. This coevolution is significant and is further evidenced through the case of the Centre for Research & Development Monitoring (ECOOM) in Flanders.

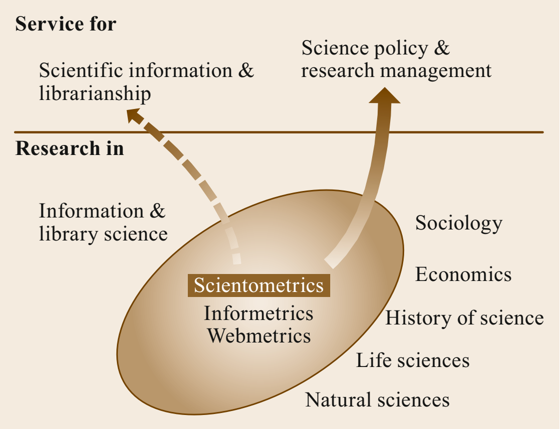


Fig. 17.1 The development and diffusion of scientometrics in the (social) sciences with application in research assessment becoming predominant (as shown by the thickness of the *arrows*)

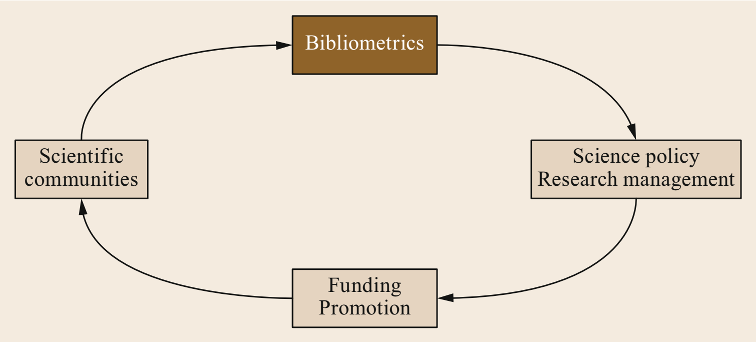


Fig. 17.2 The symbiosis of scientometrics and science policy (after [17.7]), *arrows* depicting the cycle of influence

17.2 ECOOM: An Instrument Linking Science Policy and Scientometrics in Flanders

In Flanders, the link between science policy and scientometrics is supported by and operationalized through the interuniversity consortium ECOOM, funded and monitored by the Flemish government. The precursor of ECOOM (i. e., Steunpunt O&O Statistieken or SOOS) was first created in 2001. In 2009 it was transformed into a permanent center of expertise of the Flemish government, called ECOOM (Expertise Centrum O&O Monitoring). The mission of ECOOM is to develop and maintain a performing system of science, technology and innovation indicators that provides the Flemish government with up-to-date and relevant statistical data on the science, technology and innovation performance of the Flemish region. Furthermore, ECOOM develops a portfolio of the relevant scientific research activities that support this mission. In order to fulfill this mission, ECOOM is expected to:

1. Conduct a series of targeted long-term, recurrent assignments and tasks that map, quantify and analyze the Flemish science, technology and innovation potential
2. Ascertain whether the Flemish government has the appropriate data and indicator management structure at its disposal on bibliometric, technometric, research and innovation data
3. Develop and maintain a qualified staff
4. Build, develop, extend and maintain the necessary original IT infrastructure
5. Deliver various ad hoc assignments and projects on demand for the Flemish government.

Bibliometrics is one ECOOM's core activity areas. The terms bibliometrics and scientometrics were introduced almost simultaneously by *Pritchard* [17.8] and by *Nalimov* and *Mulchenko* [17.9] in 1969. While *Pritchard* explained the term bibliometrics as "the application of mathematical and statistical methods to books and other media of communication" [17.8], *Nalimov* and *Mulchenko* defined scientometrics as "the application of those quantitative methods which are dealing with the analysis of science viewed as an information process." [17.9] According to these interpretations, scientometrics is restricted to the measurement of science communication, whereas bibliometrics is designed to deal with the more general information processes.

The sharp rise which bibliometrics has undergone since the late 1960s is reflected in remarkable academic activities. This rise is intimately connected with:

1. Advancements in information technology
2. Developments in computer science and technology
3. The worldwide availability of large bibliographic databases, serving as the groundwork for bibliometric research.

In particular, the (former) ISI databases should be mentioned in this context. The SCI, and more recently, the Web of Science, have become the generally accepted basic sources for bibliometric analysis. ECOOM has acquired access to the Thomson Reuters (now Clarivate Analytics) Web of Science (WoS) dataset, as well as to the conference proceedings data. The WoS data are available from 1980 and the conference proceedings data from 1991. In addition, ECOOM has access to an experimental version of Elsevier Scopus from 2008.

However, in the 1970s, when data collection was often still a matter of manual work, the bibliometric field was characterized by the personalities of enthusiastic researchers, whose work in the field was more of a *hobby*. Later, interdisciplinary approaches, mathematical and physical models, and sociological and psychological methods, as well as the long tradition of library science were integrated. Since the beginning of the 1980s, bibliometrics evolved into a distinct scientific discipline with a specific research profile, several subfields and the corresponding scientific communication structures. Imitating the transition from the *manufactural* form of *little science* to the *big science* of multinational research centers with enormous governmental and industrial support, scientometrics itself has claimed to have evolved from its *little* form to a *big* one with large computerized databases and with national and multinational research policy agencies as major customers.

In the 1990s, bibliometrics became a standard tool of science policy and research management. In particular, all significant compilations of science indicators heavily rely on publication and citation statistics and other more sophisticated bibliometric techniques. The IT revolution we have recently witnessed has further sped up scholarly communication. Electronic publication, open access, and the *World Wide Web* have essentially facilitated access to scientific information. ECOOM, like the Centre for Science and Technology Studies in Leiden (CWTS) or the Observatoire des Sciences et de Techniques in Paris (OST), has developed competencies in advanced bibliometric techniques and actively takes part in international, state-of-the-art research in the bibliometrics domain. But ECOOM does not limit its bibliometric activity to the large, international databases.

Given the limitations of those traditional databases to sufficiently capture and cover the bibliometric output of the social sciences and humanities, ECOOM has also developed the VABB-SHW, the Flemish Academic Bibliographic Database for the Social Sciences and Humanities (i. e., *Vlaams Academisch Bibliografisch Bestand voor de Sociale en Humane Wetenschappen*). This database captures the output of Flemish scientists from the social sciences and humanities that is not covered by WoS or Scopus. It not only includes peer-reviewed articles, but also books and monographs that have been subject to peer review. The list of journals and editors that are used for counting is determined and updated annually by the authoritative panel (*Gezaghebbende Panel*) on social sciences and humanities composed of 18 researchers in the fields of social sciences and humanities who have obtained international renown and recognition.

Since 2011, the outputs derived from the VABB-SHW, as stipulated in the Flemish university legislation on performance-based funding systems, are a parameter in the distribution of both the block grant and the dedicated research funds amongst the five Flemish universities. As such the VABB-SHW is an important addition to the Flemish performance-based research funding system. Apart from Norway [17.10], no other countries or regions have thus far succeeded in implementing a full coverage database of scholarly publications in the social sciences and humanities, making the VABB-SHW a prime international example of the innovative yet inclusive science policy pursued by the Flemish government.

We discuss the bibliometric activities of ECOOM. It should be mentioned, though, that ECOOM also contributes to the design and maintenance of technometric indicators (patent analyses), innovation indicators (based on the Frascati and Oslo manuals), indicators to measure outputs in the arts, and indicators on PhD activities in Flanders. Given this comprehensive scope, ECOOM explicitly supports the science, technology and innovation (STI) policy of the Flemish government. Not only do its various tasks and work packages support the Flemish government's recurrent needs for accurate and timely STI data and indicators, but they also provide the necessary input for dedicated policy objectives adopted by the Flemish government, i. e., the need for monitoring, benchmarking, and mapping the output of the Flemish STI system, as well as providing the necessary data and indicator support for the Flemish innovation cluster policies that have been adopted recently.

To this end, ECOOM operates as an interuniversity consortium, including the active participation of all Flemish universities: KU Leuven, UGent (Universiteit Gent), VUB (Vrije Universiteit Brussel), UA (Universiteit Antwerpen) and UHasselt (Universiteit Hasselt). It fulfills the following objectives:

- Develop and maintain a validated, robust, coherent and recurrent system of science, technology and innovation indicators
- Develop, grow and maintain the necessary and accessible IT infrastructure and cornerstone databases
- Conduct and deliver relevant and up-to-date research on the further development of STI indicator sets
- Deliver and provide the necessary indicators and statistics to the Flemish government to support its science, technology and innovation policy
- Execute a portfolio of policy-relevant studies and projects requested by the Flemish government
- Whenever necessary, reply to and advise on ad hoc questions of the Flemish government with respect to science, technology and innovation policy
- Coordinate and deliver the biennial editions of the *Flemish Indicator Book on Science, Technology and Innovation*.

17.3 ECOOM: Mapping and Benchmarking Science Activities in Flanders

Over the years, ECOOM has validated and benchmarked the bibliometric position of Flemish output in the WoS. The WoS data sources deployed by ECOOM include the Science Citation Index Expanded (SCIE), the Social Sciences Citation Index (SSCI), the Arts & Humanities Citation Index (A&HCI) and the conference proceeding citation indices. Multiple and diverse analyses are continuously executed in support of Flemish science policy.

By way of example, SCIE and the Conference Proceedings Citation Index – Science (CPCI-S) have allowed ECOOM to engage in longitudinal, citation-based analyses of Flemish scientific activity. To that end, the following indicator system was introduced in a systematic and recurrent manner. The analyses are usually based on the five so-called *relevant* or *citable* document types, namely *articles*, *letters*, *notes, proceedings* and *reviews*.

The *mean observed citation rate* (MOCR) (cf. *Braun* et al. [17.11]) provides the starting point for our analyses. MOCR is defined as the ratio of citation count (i. e., in a three-year citation window) to publication count. It reflects the factual citation impact of a country, region, institution, research group, etc. The MOCR indicator (as well as the following indicators) is used in two versions: the standard MOCR includes author self-citations, while MOCRX does not [17.12]. Normalized citation rates can then be calculated on the basis of the MOCR and the following two expected citation rates. These indicators allow us to assess the relationship between the factually achieved citation impact, the visibility/publication strategy and the subject standard:

- The journal-based mean expected citation rate (MECR)
- The subject-based field expected citation rate (FECR).

The *mean expected citation rate* (MECR) (cf. *Braun* et al. [17.11]) of a single paper is defined as the average citation rate of all citable papers (i. e., articles, letters, notes, reviews) published in the same journal in the same year. Instead of the one-year citation window to publications of the two preceding years as used in the journal citation report (JCR), a three-year citation window to one source year is used, as indicated above. For a set of papers assigned to a given country, region or institution in a given field or subfield, the indicator is the average of the individual expected citation rates over the whole set. Analogously to the previous indicator, the FECR of a single paper is defined as the average citation rate of all papers published in the same subject in the same year.

Those indicators form an indicator triplet that should best be considered and interpreted together. Their mathematical relation reveals details about publication strategy and factual impact with respect to what should be expected on the basis of the publications' subject. Several *configurations* are possible, for instance: MOCR > MECR > FECR, which reflects the most favorable situation, and means that the author under study publishes on average in journals with higher-than-discipline standard and receives even more citations (on an average) than the standard set by the journals in which the papers are published. MECR > MOCR > FECR means that the latter standard is not reached and, for instance, FECR > MOCR > MECR means that the researcher achieved a higher citation impact than expected on the basis of the journals in which he/she has published but these journals do not, on an average, belong to the top journals in their discipline. We just mention in passing that these relations can also be expressed numerically by forming the ratios MOCR/MECR, MOCR/FECR and MECR/FECR with the neutral value 1.0.

Braun and *Glänzel* [17.13] have introduced the foundations of the use of those indicators in the context of measuring national publication strategy, and they have been used in a multitude of quantitative studies of national, regional and institutional research assessment [17.6]. Versions of these indicators are used also at CWTS in Leiden [17.14].

In addition to these standard indicators we provide a more versatile, however also more complex, tool for measuring research performance as reflected by citation impact. The distribution of number and share of papers in so-called performance classes representing moderately and highly cited papers according to a four-class scheme of the self-adjusting method of characteristic scores and scales (CSS) is compared with the expectation based on the world standard. These scores and scales are obtained from iteratively truncating samples at their mean value and recalculating the mean of the truncated sample until the procedure is stopped or no new scores are generated. Usually three scores are sufficient, where the first one is identical to the mean value of the reference population. The resulting four classes are obtained by the intervals defined by adjoining scores (Chap. 13 by *Glänzel* et al. in this volume, [17.15, 16]). This method is a real alternative to the application of percentiles, but has two important advantages: 1) CSS is not biased by ties in the underlying citation ranking, and 2) CSS scores are self-adjusting and thus not defined on arbitrary preset values.

The four classes stand for:

1. Poorly cited papers
2. Fairly cited papers
3. Remarkably cited papers
4. Outstandingly cited papers.

Papers in classes 3 and 4 can be considered highly cited. CSS provides robust classes in terms of their insensitivity to publication year, citation windows and subject. Although CSS is not directly linked to percentiles, the standard distribution of papers over classes is about 70% (1), 21% (2), 6–7% (3) and 2–3% (4). The deviations of the institution's profile from the standard or from that of another institution provide a multifaceted picture of its citation impact. An institution's share in certain classes might be higher or lower than, or equal to, the corresponding standard and its profile might thus follow the above-mentioned reference standard or be more or less polarized than the standard or more skewed towards poorly or highly cited papers, respectively. An institution might have more highly cited papers than expected and at the same time less poorly cited papers than expected, but it might also be the other way around and thus have both more poorly and highly cited papers than the reference standard.

By way of example, an overview of some recent analyses is provided in the subsequent figures and tables. This overview gives a flavor of the richness and diversity of scientometric insight ECOOM provides to the Flemish government in support of its science policy. Figure 17.3 provides an overview of the Flemish output (SCIE + CPCI-S) per 10 000 inhabitants over the period 2002–2013 compared to 12 benchmark countries selected by the Flemish government given their relevance to Flemish science policy. Given the longitudinal perspective taken, it provides the Flemish government with a detailed overview of the evolution of Flemish scientific output. This output is further broken down into institutional categories (universities, strategic research centers, hospitals, companies) and into scientific disciplines and subdisciplines based on the Leuven–Budapest subject classification scheme as developed by *Glänzel* and *Schubert* [17.17] in 2003. Those recurrent analyses allow for a detailed and systemic understanding of Flemish research output and productivity. A standard set of 12 countries is used in those analyses. In addition to those 12 *standard* benchmark countries, comparisons with China, Japan and the US are also made on a regular basis.

Figure 17.4 provides an overview of the relative citation frequency of Flemish science and international

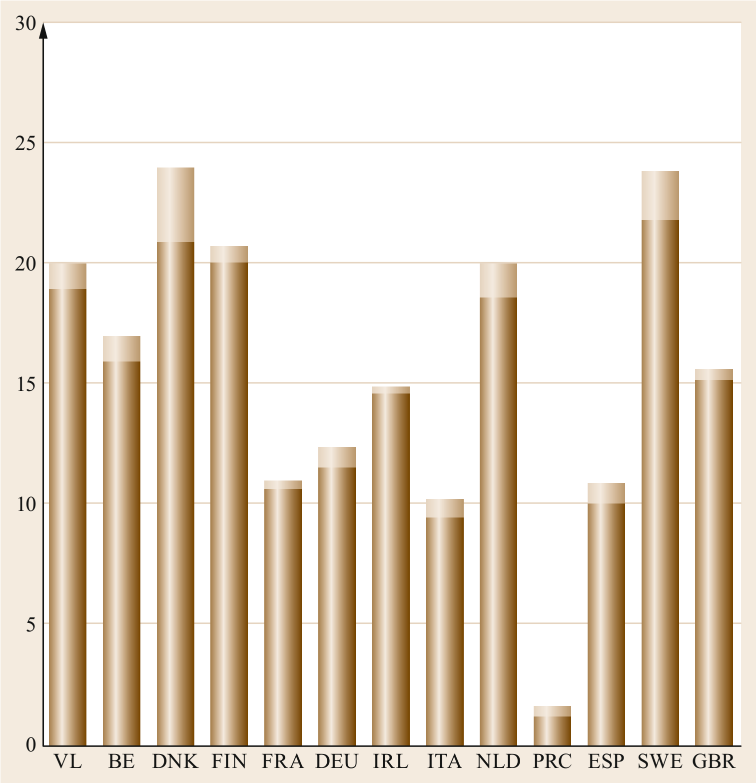


Fig. 17.3 Overview of the Flemish output (SCIE + CPCI-S) per 10 000 inhabitants over the period 2002–2007 (*dark bars*) and 2008–2013 (*light bars*), compared to 12 benchmark countries

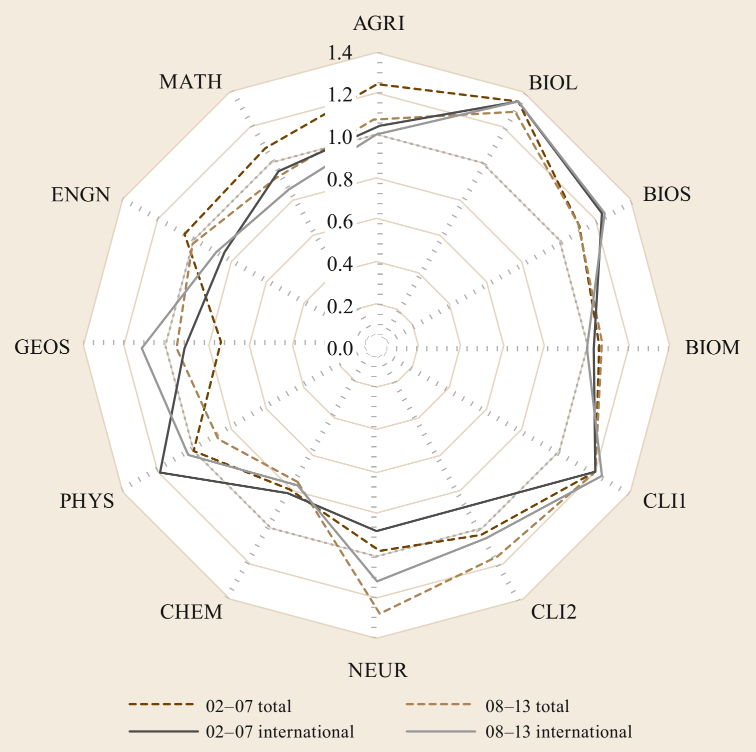


Fig. 17.4 Overview of relative citation frequency and international collaboration across 12 disciplines of the sciences in Flanders

collaboration across 12 disciplines of the sciences. Such citation analyses and collaboration analyses are another benchmark ECOOM computes for the Flemish government. This allows the Flemish government to assess the international visibility of Flemish science and its connectedness to the rest of the world.

Table 17.1 maps the bibliometric performance of a Flemish strategic research center using the *standard* ECOOM indicator set. This type of analysis is done regularly for all Flemish strategic research centers and allows the Flemish government to assess and benchmark their performance. Table 17.2 maps the bibliometric performance of a Flemish strategic research center using the ECOOM CSS approach, highlighting the citedness of the research output of the center under examination.

Those analyses are illustrative of the scientometric/bibliometric stance taken by ECOOM in its various tasks and activities to support Flemish science policy. Every two years, they are bundled into a new version of the Flemish STI indicator book. However, they do not only provide *factual*, quantitative input to the Flemish government as to the status of the Flemish science landscape and the performance of its actors. They are also used as one of the underpinning elements for expert assessments and informed peer reviews of the perfor-

Table 17.1 Bibliometric performance of a Flemish strategic research center using the *standard* ECOOM indicator set

Period	Publications	Citations	Self-citation	MOCR	RCR	NMCR	NMCR/RCR
2000–2004	1575	22 883	19.7%	14.53	1.25	1.98	1.58
2005–2009	2010	32 797	20.8%	16.32	1.22	2.10	1.73
2010–2012	1659	36 175	21.3%	21.81	1.40	2.80	2.00
2000–2012	5244	91 855	20.7%	17.52	1.29	2.29	1.77

Table 17.2 Bibliometric performance of a Flemish strategic research center using the ECOOM CSS approach

Period	Publications	Class 1	Class 2	Class 3	Class 4
2000–2004	1575	44.4%	32.6%	14.3%	8.6%
2005–2009	2010	41.6%	32.1%	18.3%	7.9%
2010–2012	1659	38.1%	32.2%	17.2%	12.5%
2000–2012	5244	41.4%	32.3%	16.8%	9.6%

mance of the various science organizations in Flanders. The main focus of those expert and peer-review exercises concerns the five-yearly reviews and regular monitoring of the strategic research centers (IMEC-iMinds (Interuniversitair Micro-Elektronica Centrum), VIB (Vlaams Instituut voor Biotechnologie), Flanders' Make and VITO (Vlaamse Instelling voor Technologisch Onderzoek)), supporting the composition of the scientific panels of the fund for scientific research Flanders (FWO), input to the multiannual expert evaluation of the extraordinary research fund (BOF) that provides lump-sum support for fundamental research at the five Flemish universities, input to the multiannual expert evaluation of FWO, input to the expert evaluation of various research institutes and initiatives supported by the Flemish government, etc. Hence, the ECOOM data sources and indicators are used extensively as an input to the research assessment endeavors that have become a central dimension of Flemish science policy.

In addition, ECOOM is asked regularly to engage in so-called domain studies. Those are forward-looking studies executed by the Flemish government to identify future areas or *domains* of science investment. Based on its bibliometric data sources and indicator base, ECOOM provides relevant bibliometric overviews and insights to such studies. In the past, domain studies were executed for such diverse fields as nanotechnology, translational medicine, plant biotechnology, stem cell research, bioinformatics, advanced materials, renewable energy, and the main Horizon 2020 key enabling technologies (KETs) as identified by the European Commission (https://ec.europa.eu/programmes/horizon2020/en/area/key-enabling-technologies).

More recently, ECOOM has been involved in providing input into such areas as industry 4.0, sustainable chemistry and circular economy. The bibliometric input to those domain studies provides the Flemish government with an in-depth insight into the structure and global evolution of the science base of the respective *domains*, into the main actors and their accomplishments, and into the relative position and accomplishment of Flemish research in the respective global contexts. In addition to the data and indicator sources described above, the use of advanced text mining techniques and search algorithm designs is high on the ECOOM research agenda as they allow for more detailed, fine grained and insightful analyses of the domain under study and the various cognitive streams of inquiry they consist of [17.18].

17.4 ECOOM: Input for Funding Formulas of Science Activities in Flanders

The publication and citation analyses of ECOOM are not only used for the mapping and assessment exercises as described previously. They are also an integral part of three major performance-based research funding systems deployed by the Flemish government. Bibliometric data are used as a significant component of three formulas that distribute lump-sum funding amongst the five Flemish universities. They are:

1. The block grant (approximately one billion Euro distributed annually amongst the five Flemish universities, aimed at capacity building in teaching and research)
2. The extraordinary research fund (BOF, approximately 160 million Euro distributed annually amongst the five Flemish universities to stimulate research excellence on the basis of peer-reviewed professorships and projects that are awarded by each university within the confines of the lump sum received)
3. The industrial research fund (IOF, approximately 30 million Euro distributed annually amongst the five Flemish universities to stimulate excellence in technology transfer on the basis of expert-reviewed innovation mandates and projects that are awarded by each university within the confines of the lump sum received).

Those funding formulas distribute the money as lump sums between the universities. The allocation within the universities is based on internal allocation rules (block grant), traditional scientific peer review (BOF) or expert review also involving representatives with industrial research and development and venture capital backgrounds (IOF). Each funding formula is built on a diversity of indicators, of which the bibliometric ones are only one set, albeit a significant one. The block grant formula further includes various educational indicators; the BOF formula also includes indicators on PhD activities as well as faculty mobility and diversity; while the IOF formula includes such parameters as patenting and spin-off activity, income from industrial collaborations and income from presence in the European framework programs. As the bibliometric component in the BOF formula is the basic one that is used across the three funding formulas, a brief overview of its design and evolution is provided here (for an overview of the origins of the BOF funding formula, reference is made to *Debackere* and *Glänzel* [17.5]).

The evolution of the BOF funding formula reads as follows:

- Phase I (prior to 2003):
  - Funding allocated on the basis of parameters based on the amount of block grant funding, student numbers and PhDs awarded. PhDs were the only research-related parameter in the allocation rule.
- Phase II (2003–2007):
  - Inclusion of bibliometric data, accounting for up to 30% of the allocation formula. Focus was on the WoS-SCIE data source. Data used: first-order publication counts (document types: article, letter, review, note) and citation counts over a ten-year time window (funding allocation in period t is based on publications appearing in period $[t-11, t-2]$ and on citations to those publications over the same time window). Both are weighted equally. The main effort consisted of cleaning and assigning publications to Flemish academic institutions, which were validated and corrected by the respective universities themselves.
- Phase III (2008–2011):
  - Extension of the bibliometric data, including publications originating from the WoS-SSCI and the WoS-A&HCI. This was a first step towards a full(er) inclusion of the research output in the social sciences and arts and humanities
  - Conference proceedings papers are also included (Conference Proceedings Citation Index – Science (CPCI-S) and Conference Proceedings Citation Index – Social Science & Humanities (CPCI-SSH)).
  - In addition, for those publications that appear in journals with an impact factor, the publications are weighted with the JCR impact factor
  - Evolution of the weight of the bibliometric component to 36% of the total funding formula by 2012
  - Segmentation into 16 major disciplines: 13 SCIE-based and three SSCI & AHCI-based (based on the scheme developed by *Glänzel* and *Schubert* [17.17])
  - Sophistication of the allocation rule increases considerably during this phase.

  During this third phase, the share of each university in the bibliometric part of the funding formula is computed as:

Definition of Variables 17.1

Formula deriving university shares of WoS outputs across discipline-based and impact-weighted publication counts and across citation counts over the period 2008–2011

$$
\begin{aligned}
B_\mathrm{u} = {} & \frac{gB_1}{gB}(g\mathrm{PSSI} \times \mathrm{BSSI_u}) \\
& + \frac{gB_1}{gB}(g\mathrm{PAH} \times \mathrm{BAH_u}) \\
& + \frac{gB_1}{gB}(g\mathrm{PR} \times \mathrm{BPR_u}) \\
& + \frac{gB_1}{gB}(g\mathrm{PSS} \times \mathrm{BSS_u}) + \frac{gB_2}{gB}\frac{C_\mathrm{u}}{\sum_i (C_i)}\,.
\end{aligned}
$$

B_u	share in bibliometric output B of each university u
gB_1	weight of the publication component in the funding formula
gB_2	weight of the citation component in the funding formula
gB	$gB_1 + gB_2$.

Publication counts:

- $P =$ sum of publications across universities, weighted 0.5 for proceedings and 1 for journal articles—$P = \mathrm{PSSI} + \mathrm{PAH} + 0.5\mathrm{PR} + \mathrm{PSS}$
- $\mathrm{PSSI_u} =$ total number of SCIE or SSCI publications with impact factor for university u
- $\mathrm{PSSI} = \sum_i(\mathrm{PSSI}_i)$, sum of SCIE and SSCI publications with impact factor across all universities
- $\mathrm{PAH_u} =$ total number of A&HCI publications for university u
- $\mathrm{PAH} = \sum_i(\mathrm{PAH}_i)$, sum of A&HCI publications across all universities
- $\mathrm{PR_u} =$ total number of proceedings papers for university u
- $\mathrm{PR} = \sum_i(\mathrm{PR}i)$, sum of proceedings papers across all universities
- $\mathrm{PSS_u} =$ total number of SCIE or SSCI publications without impact factor for university u
- $\mathrm{PSS} = \sum_i(\mathrm{PSS}_i)$, sum of SCIE and SSCI publications without impact factor across all universities
  - $g\mathrm{PSSI} = \mathrm{PSSI}/P$
  - $g\mathrm{PAH} = \mathrm{PAH}/P$
  - $g\mathrm{PR} = 0.5\mathrm{PR}/P$
  - $g\mathrm{PSS} = \mathrm{PSS}/P$
  - Sums i are across all universities.
- $\mathrm{BSSI_u}$, $\mathrm{BAH_u}$, $\mathrm{BPR_u}$, $\mathrm{BSS_u}$ then represent the (where possible, discipline-based, impact-weighted) publication shares of university u for each of the four WoS sources.

Citation counts:

- C_u = total number of citations C to all publications of university u
- $\sum_i(C_i)$ = total number of citations C to all publications of all universities.

- Phase IV (post 2011):

  - The presence of arts and humanities and social sciences in the allocation rule is further developed and increased through the creation of a unique flemish academic bibliography (VABB-SHW), taking into account a broader (non-WoS-based) set of relevant publications from journals, books, monographs, proceedings, catalogues, ..., inspired by the Norwegian example
  - An authorative panel (*Gezaghebbende panel*) composed of 18 renowned scientists from the social sciences and humanities selects and approves the source materials included in the new database
  - Every 3 years, an evaluation panel consisting of minimum 5 researchers in the social sciences and humanities disciplines will assess the quality and validity of the VABB. These researchers should be familiar with the flemish research activity in the respective disciplines, while not being active in Belgium at the time of the evaluation.
  - Finally, the sophistication of the weighting formula further increases. The 16 major disciplines deployed in phase III are extended to 68 sub-disciplines, still according to the subject classification scheme developed by *Glänzel* and *Schubert* [17.17]. Instead of weighting the publications by their JCR impact factor (when available), journals are now classified into 20 5% intervals, ranking them from high to low based on their JCR impact factor (of course, for those disciplines where journal impact factors are available). The 20 intervals are then weighted equally across the 68 sub-disciplines, thus eliminating the cross-disciplinary distortions that originated from using the absolute impact factor values as weights (Table 17.3).

Although the Belgian Federal Science Policy Office (Belspo) regularly maps the aggregated Belgian publication output, no approach or setup equivalent to ECOOM exists in Wallonia at the moment.

To conclude, over the years, the bibliometric funding formula in Flanders has become more sophisticated. This sophistication originated as a result of a dialog between science policy and scientific community. Two major advances brought about by this dialog are:

1. The better recognition and more exhaustive inclusion of the social sciences and humanities in the funding formula
2. The adjustment of the publication weights taking into account the differences in citation patterns and subsequently impact factors across disciplines.

For sure, the formula thus became more complex, but at the same time it has gained in fairness across the various disciplines present at the five Flemish universities. In addition, weighting the publications this way is expected to reduce excessive publication behavior and hence to avoid large numbers of publications in low-impact journals. *Linda Butler* observed those excessive publication behavior effects in Australia [17.19].

The publication and citation components as computed for the BOF funding rule spill over into the publication and citation components of both the block grant and the industrial research fund (IOF) funding rules (although with different weights, given the different finality of those two other funding instruments).

Finally, it should be noted that data retrieval, validation and computation are done with absolute transparency. Each university has access to all data (also the ones of the other universities) and validates and corrects them before they are used in the funding formula, the computational outcome of which is also first verified by all institutions before it is applied to distribute the funding itself. As a consequence, fairness and transparency have been important drivers of the continuous development and adjustment of the funding formulas and their operationalization.

Table 17.3 Weights for the 20 Impact-Factor-based 5%-intervals ranked in descending order for the BOF formula

Interval	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Weight	10	6	3	2	1	1	1	1	1	1	1	1	1	1	0.5	0.5	0.1	0.1	0.1	0.1

17.5 ECOOM: No Data and No Indicators Without a Solid IT Backbone

The key to—and one of the main pillars of—successfully delivering services to science policy stakeholders is the development, deployment and maintenance of a high-performing IT platform that satisfies the specific requirements of a research and service environment. Its main characteristics at ECOOM are:

- Ease of processing: The calculation of standard indicators on a set of publications must be a straightforward exercise. Also, research and services must not suffer from limitations in computational power. ECOOM therefore developed its own ASCII-based data format that proves to be very fast for processing and calculations for default, recurring tasks. For complex tasks with high computational demands, ECOOM has developed an extensive toolbox for using dedicated servers for parallel computing and hyper-threading or for using the shared memory cluster available in the universities' high performance computing centers or (when massive storage is needed) to access the Elastic MapReduce (EMR) service offered by Amazon web services.
- Large storage capacity: The data sources used for bibliometric research are to be stored locally. Given the size of these databases (several terabytes), an extensive storage capacity is needed. Storage alone is not sufficient, given the tremendous efforts put into preparing data, so a reliable backup policy is mandatory
- Fast retrieval: Getting the right data or information out of an archive or database is even more important than entering the data in the first place. Entering the data in the general Microsoft SQL Server database platform set up at ECOOM allows for fast retrieval of publications or counts of journals, countries, and fields. This is extended by NoSQL framework consisting of a Lucene text-index for text-based retrieval; a Neo4J graph database for authors and institutions; and finally an extensible markup language (XML)-based document store
- Accessible: All of ECOOM's partners need access to the bibliographic databases. This access is provided by the relational database. However, this accessibility comes with a security issue. Therefore, we have chosen to implement this on a dedicated Hyper-V virtualized environment with all necessary security precautions.
- Reliability: This is provided at different levels through the general ECOOM IT infrastructure, with a reliable backup policy implemented, hardware with all necessary redundancy built in, and mirror databases on virtual servers.
- Data quality: Indicators are only as valid as the quality of the underlying data permits. Therefore, the platform must enable us to ensure the highest possible quality of data for the tasks at hand. As research is done on different units of analysis and with different objectives, the procedures for data cleaning must be flexible and scalable.

The data on publications and citations—needed for the bibliometric services we provide—are gathered from two data providers: Clarivate Analytics (previously Thomson Reuters) and Elsevier. Currently, Clarivate provides us with the data available in the Web of Science, consisting of the following datasets: science citation index expanded, social science citation index, arts and humanities citation index, the two conference proceedings citation indexes, and the emerging sources citation index (ESCI). The journal citation report is also provided.

After ECOOM has explored the two editions, *Conference Proceedings Citation Index – Science (CPCI-S)*, and *Conference Proceedings Citation Index-Social Sciences and Humanities (CPCI-SSH)* of Thomson Reuters' (now Clarivate) proceedings database, a new, essential bibliographic source is now available. The Book Citation Index (BKCI) with a focus on technical sciences, social sciences, and humanities provides citations from journals, proceedings and other books. This new index is operational at the chapter level, as well as at the complete book level. Unlike the journal and proceedings citation indexes, several problems emerge in the context of subject classification and citation processes.

Establishing a categorical structure for books will become a new challenge, since this structure might considerably differ from the one used for journals and proceedings. A second issue refers to the aging of information. Aging, i. e., information use measured by citations, might be longer than in the case of periodicals. In this context, the issue of different editions might also emerge. The objective of ECOOM is to establish relevant and sensitive, but also robust baselines. Such baselines are indispensable for benchmarking and evaluative studies.

The exploitation of the opportunities offered by the second, large abstract and citation database—Elsevier's Scopus—forms an important task in the extension and validation of bibliometric studies. Although both WoS and Scopus have, in principle, similar features, their internal structure does essentially differ. Scopus subject classification based on journal assignment, as well as unification of author and institutions, considerably dif-

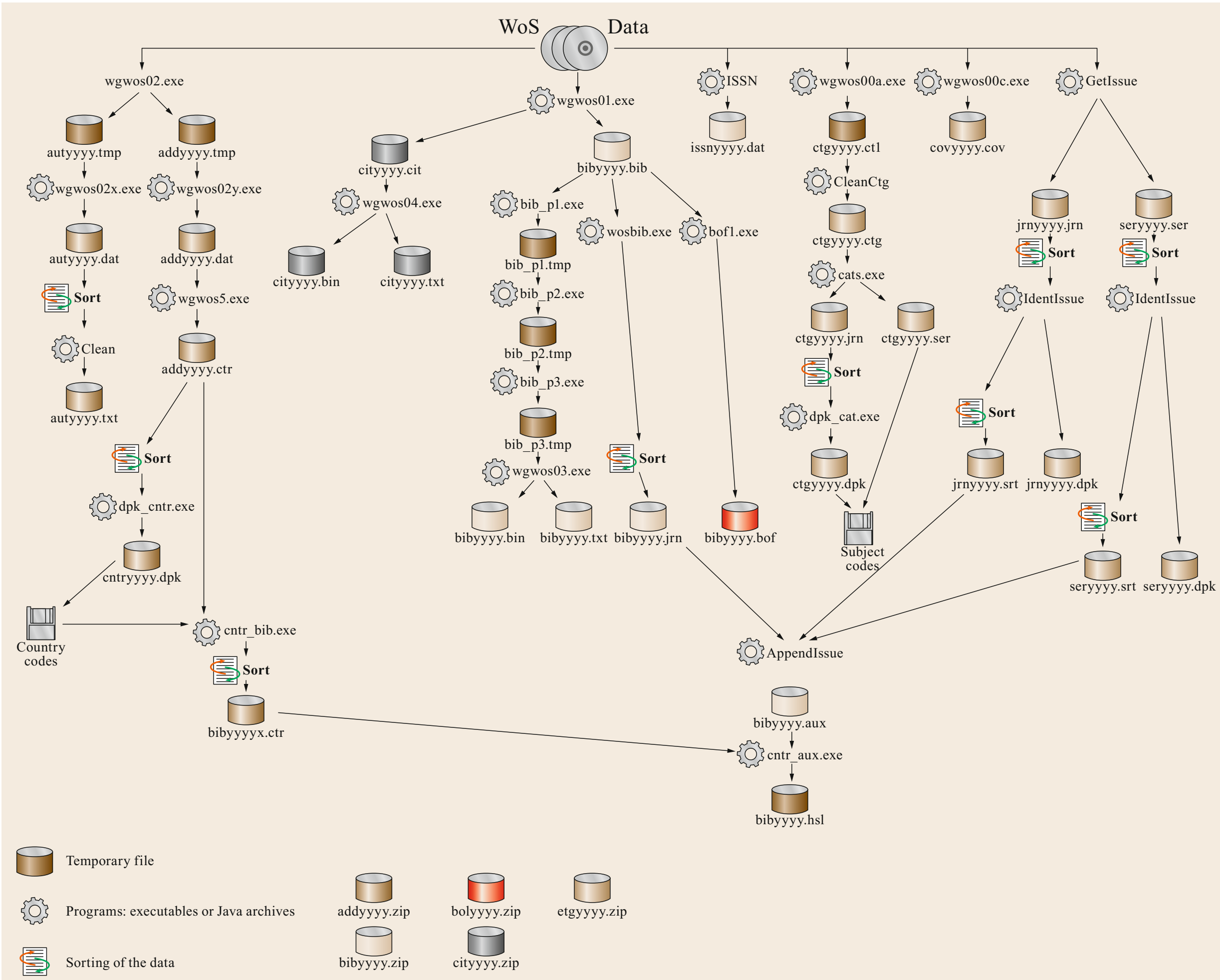


Fig. 17.5 The process of data cleaning and preprocessing for the relational database and the modular indicator-file system

fer from the classification and assignment provided by the Web of Science. The elaboration of a possible concordance between the two systems promises to become an important step in the improvement of the coverage and the reliability of bibliometric data sources.

The existing web interfaces to these databases do not meet the requirements ECOOM needs to perform its tasks. That is why, to be able to implement our own solutions, we receive the entire contents of these databases. Clarivate provides its yearly data volumes of flat text files in a tagged format and the complete custom-generated data including the book citation index in a XML format. Elsevier delivers the Scopus data as XML files.

These raw data are not in a format suitable for indicator calculations, so extensive processing needs to be done. For both datasets a three-step approach is taken. First, elaborate preprocessing is needed. Data files are checked for data integrity or file corruption, and corrections are made based on supplied information.

Next, effective processing is started. Procedures have been developed and improved over the years, using JAVA, Pascal, and AWK. As the provided data format changes, or new information becomes available, or new insights in the data are gained, adaptations need to be made to these procedures. Data are extracted from the source files, and are transformed and combined. Due to the huge amount of data to be processed (in the terabyte order of magnitude), many of the algorithms are adapted over time to allow parallel processing on our multicore servers by using the hyper-threading capabilities of the JAVA programming language or to distributed implementations on EMR.

For the deployment of the data, two paths are followed. For the recurring tasks, using the custom-built

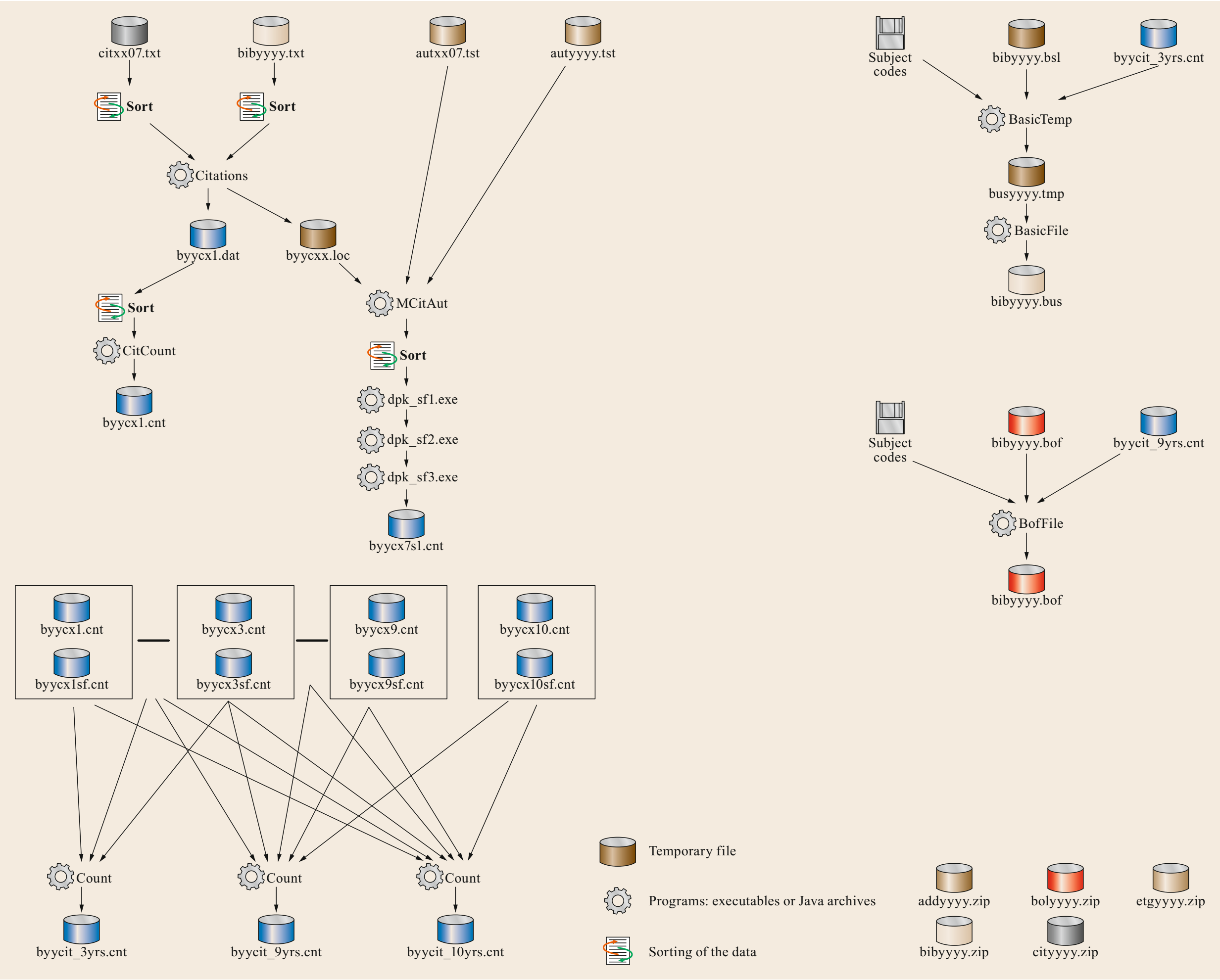


Fig. 17.5 (continued)

flat-file database is the optimal data structure. The procedures on the flat-file database are implemented in JAVA and Pascal. For other tasks, the data is uploaded into a Microsoft SQL Server relational database or the other data stores. For the XML data from WoS and Scopus databases additional steps are required using XQuery statements with FLOWR and XPath expressions to extract the required information.

Once the data are available in a relational database, materialized views and indexes are generated to meet the performance needs. Additionally, this step often requires changes to the database, the operating system, and sometimes even the hardware of the server the database runs on.

A copy of the relational database structure of the WoS data is moved to the dedicated server to ensure accessibility for all of ECOOM's consortium partners. All flat files and the database are backed up using the Tivoli TSM service provided by the central KU Leuven IT department. For the programming code, a SVN repository is deployed to ensure reliable storage and version control.

Processing and retrieval are further enhanced by the application of newly developed techniques based on hybrid combinations of text mining and citations. They also proved to be the most promising methods in the analysis of the epistemological structure of science [17.20–23]. Text mining can aid the summarization, categorization, and interpretation of large sets of documents and the analysis of their dynamics. It can provide powerful tools for the mapping of science. Citation links, in turn, provide a *natural* measure of the relatedness of documents. ECOOM has developed techniques for combining the advantages of both approaches and for compensating their shortcomings at the same time [17.24].

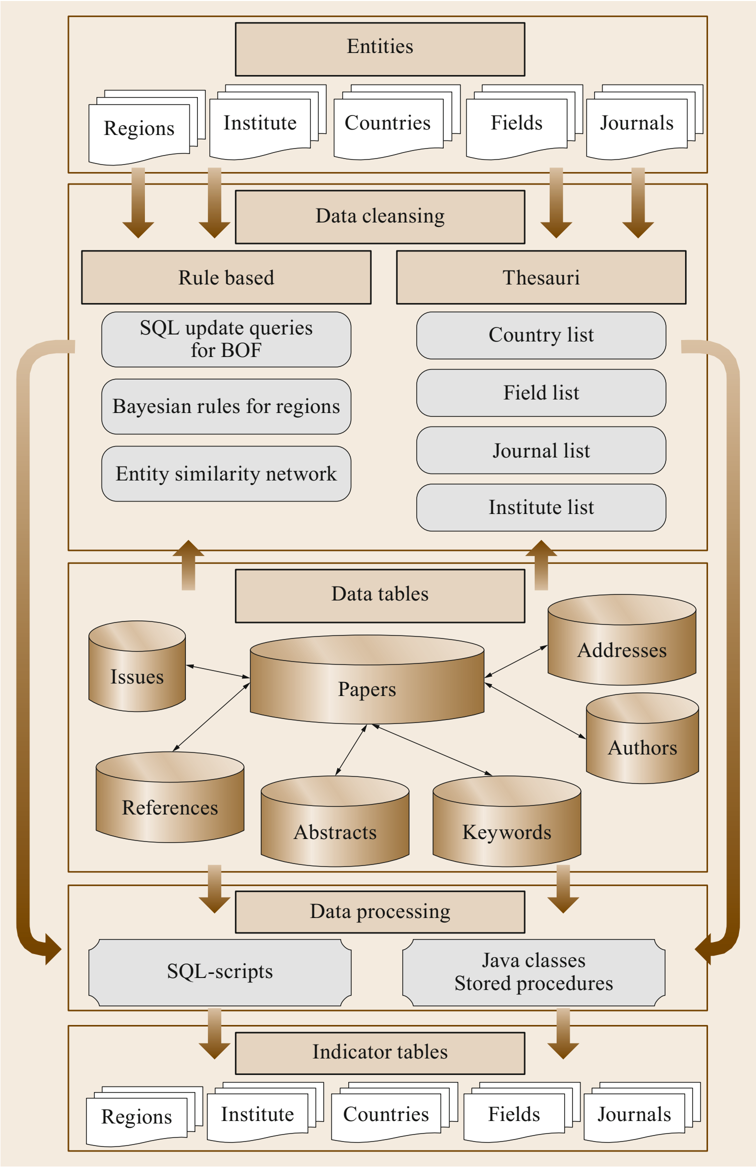


Fig. 17.6 Structure of the relational database created from the abstract and citation databases

The research component in ECOOM's bibliometric service task aims at achieving two major objectives: 1) improving subject classification systems for bibliographic databases, such as the one developed by ECOOM [17.17] and, 2) the structural analysis of the document space. The latter approach has two important applications: the mapping and visualization of the structure of science and its dynamics, and the detection of emerging research topics. The application of the hybrid clustering of the document space will thus be done both globally (i. e., based on the complete database and/or major subject areas) and locally, that is, within narrower disciplines. While the first—the global approach—requires not only efficient dimensionality reduction algorithms, like latent semantic indexing (LSI) or singular value decomposition (SVD), but also improved feature creation based on noun phrases by the application of natural language processing and intelligent matching procedures like locality sensitive hashing. The structural topic analysis and the dynamic analysis—which is required for emergence detection at the local level—are faced with a different problem. Topic description cannot usually be based on the textual component alone, since the same vocabulary is commonly used within the often narrow discipline. In previous studies, ECOOM has shown that so-called *core documents*, i. e., frequently *and* strongly interlinked items, can readily be used for the description, characterization and labeling of topics at the local level [17.18], and also for dimensionality reduction on a large scale.

ECOOM continuously improves its data capabilities. The direct clustering of the document space, based on hybrid similarities, is now supplemented by what we will call second-order similarities. The idea is adopted from social network studies, where a strong correlation of mediate similarities is used to detect profile congruence of individual documents. In other words, instead of the similarities of direct links between documents, the correlation of their similarity profiles is used as the basis for clustering. This provides an additional viewpoint on what forms the basis of the cognitive environment of publications.

The issue of subject classification has already been mentioned; in addition, the combined text- and citation-based methods provide techniques for the delineation of (interdisciplinary) fields and topics, which the classification scheme does not immediately provide, such as stem-cell research or bioinformatics. Finally, the detection of emerging fields also provides the groundwork for the identification of the main players, and for evaluative and comparative bibliometric studies of these fields.

Data quality is related to the different services required by public agencies requesting ECOOM input at the national (OECD, Eurostat), regional (Flemish government, agencies like EWI, VLAIO and VARIO), institutional (BOF) or individual researcher (FWO) level. Unfortunately, author and institution names are not unique in the databases provided. Institutes or authors are recorded in the database as they appear on the publication. The use of multiple names, spelling variations, possible errors, or the use of different initials, make it hard to gather all information surrounding a certain entity of analysis. To solve this, record linkage techniques have to be applied to the data. This encompasses data cleaning, unification, duplication, and enriching of the data. Since a 0% fault-tolerance is applicable to some of the tasks (e. g., BOF and FWO), large portions of this work have to be done manually. We are therefore continuously looking into technologies to speed up this process.

The direct result of speeding up the rate at which data can be cleaned is a faster response rate to various policy requests. Yet more influential is the use of the time that becomes available. This time needs to be spent on increasing data quality.

Finally, the web-based data sources are also analyzed for their possible use in evaluative bibliometrics. Opportunities and limitations of these sources will be compared with those of traditional abstract and citation databases. Repositories like ResearchGate or Google Scholar and its various derivatives thus offer a main subject of analysis. The major IT backbones for the bibliometric activity are briefly summarized in Figs. 17.5 and 17.6.

17.6 Insights Obtained

Taking into account the rich indicator- and case-based evidence stemming from 16 years of ECOOM bibliometric experience in Flanders, the symbiotic role of scientometrics in shaping science policy and of science policy in setting the scientometric research and activity agenda has been illustrated. Two important dimensions of the *scientometrics–science policy* symbiosis can be recognized. One is its ex ante role. The other one its ex post role.

The ex ante role of scientometrics has become clear in various ECOOM cases and activities, such as:

- Informing the genesis of novel science policy instruments and areas of attention, e. g., the contribution to various domain studies on emerging fields of science that provide informed input to policymakers and create new supportive arrangements for such areas
- Supporting the design and evolution of performance-based research funding formulas like the BOF funding formula.

Informing and supporting science policy in turn generates new demands on the quality of scientometric data and indicators as, for instance, in the case of understanding the nature of scientific excellence. To this end, advanced research into the development of a state-of-the-art indicator base is absolutely necessary. However, this is not sufficient. In order to be productive and responsive, a well-performing and accessible IT system supporting the scientometric research and service tasks is required as well.

The ex post role of scientometrics also became clear in various ECOOM cases and activities, such as:

- Mapping cognitive structures, actor connectivity and institutional performance in the Flemish academic system
- Assessing the multilevel scientific performance of institutional actors in the Flemish science system
- Monitoring multiannual strategic plans and accompanying funding schemes for various actors in the

Flemish science system, based on the productivity and visibility of their most recent research activities.

This combination of ex ante and ex post roles leads to an intense symbiosis between developments in scientometrics and the design of science policy trajectories and instruments. The evolution, optimization and rejuvenation of science policy instruments are enabled and facilitated as a result. This evolution, optimization and rejuvenation of science policy instruments continuously necessitates the field of scientometrics to come up with novel, better, more relevant, more valid, more robust methods and indicators. The widespread advent and development of performance-based funding models for research organizations further underpins the above considerations [17.25, 26].

Although well entrenched in the Flemish science policy scene, the use of quantitative methods and metrics-based funding models and policy instruments will never be uncontested. Both policy makers and scientometricians therefore need to carefully understand their limitations as well as potential abuses. One of the basic principles adhered to in the ECOOM context is the complete openness of data and indicators amongst the actors involved, as described. Also, the design and use of the funding formulas and other policy instruments never is a top-down government decision, but always involves the participation of the actors involved and an intensive, exhaustive dialog between the science policy scene and the scientific community. Not without reason has ECOOM been constructed, deployed and funded as an interuniversity consortium and not as a government agency. This approach highlights the basic trust and dialog that should prevail in developing and growing a symbiotic relationship between science policy and scientometrics. And even with this zero-order principle in place, all stakeholders have to be aware of distorted behaviors that may result from the policy use and misuse of bibliometric data.

One issue concerns the changes in the publication, citation and collaboration behavior of scientists (both positive and negative) that the consistent policy use of bibliometric indicators might potentially induce. Studies on the problem choice behavior of academic scientists have revealed that both cognitive and social influences determine the manner in which scientists go about choosing the problems they work on [17.27]. Hence, the issue always has to be raised as to what extent the policy use of bibliometrics may or may not affect this behavior.

The problem of the inappropriate use of scientometrics or bibliometrics ranges from uninformed use, through selecting and collecting *most advantageous* indicators, to the obvious and deliberate misuse of data. Uninformed use and misuse are not always beyond the responsibility of bibliometricians. Unfortunately, bibliometricians do not always resist the temptation to follow popular, even populist trends in order to meet the expectations of the customers. Clearly, any kind of uninformed use or misuse of bibliometric results involves the danger of bringing bibliometric research itself into disrepute [17.28].

Uniformed use consists of:

- Incorrect presentation or interpretation of bibliometric indicators or their use in an inappropriate context caused by insufficient knowledge of methodology, background and data sources
- Generalization (*induction*) of special cases or of results obtained at lower levels of aggregation.

Misuse consists of:

- Intentionally incorrect presentation or interpretation of bibliometric indicators or their deliberate use in inappropriate contexts
- Tendentious application of biases
- Tendentious choice of (incompatible) indicators.

But even correct use might have undesired consequences. For instance, reinterpreting underlying contexts such as the notion of *citation* can show author self-citations in an unfavorable light. Authors might thus be urged to avoid self-citations—a clear intervention into the mechanism of scientific communication, which is not desirable if deployed in an uninformed, linear way. Less obvious repercussions may be observed when bibliometric tools are used in decision-making in science policy and research management and the scientific community recognizes the feedback in terms of their funding. This is one of the criticisms of performance-based funding formulas that we also encounter in Flanders. Hence the need to revisit and to adjust the funding formula regularly, as new and deeper insights on its positive, but also negative, effects emerge. *Butler* [17.19], for instance, has shown, based on the example of Australia, what may happen when funding is linked to publication counts. She found that the publications component of the composite index used at that time in Australia has stimulated an increased publication activity in the lower-impact journals. It is worth mentioning here that, in Australia, journal rankings were designed, published, used for evaluation purposes, and then retracted. See for instance the Australian ERA that was dismissed for being ill conceived and for compelling researchers to publish in selected venues rather than in venues suitable to their research. A revisit of Fig. 17.2 brings us back to the possible feedback loops of the policy use of bibliometrics on the scientific community. This feedback may, as we know in the meantime, both have positive and negative effects.

Possible positive effects are:

- Scientists might recognize that scientific collaboration and publishing in high-impact or even top journals pays
- Their publication activity might be stimulated.

Possible negative effects are:

- Exaggerated collaboration and even trends towards hyperauthorship, inflating publication output by splitting up publications into sequences (or *salami slicing*), inflating citation impact by self-citations and forming citation cliques, etc.
- Trends towards replacing quality and recognition by visibility at any price or towards preferring journals as publication channels in social sciences and humanities.

Fortunately, the increasing sophistication of bibliometric research provides us with the insights and the tools required to cope with such use, abuse, and positive and negative effects. Normalization and standardization are the mathematical avenues that can be pursued to explore and to remedy them. The dialog and transparency between the scientific community and the science policy scene as it is operated in Flanders through the creation of ECOOM is yet another, organizational mechanism to cope with the aforementioned effects. And, finally, we should never forget to fully use the rich complementarity that exists between peer review and the metrics toolboxes that have been developed [17.29]. Hence, a rich and varied array of approaches can be mastered and deployed to monitor the symbiosis between scientometrics and science policy. ECOOM hopefully provides an inspiring and interesting case in this respect.

References

17.1 J. de Solla Price: *Big Science, Little Science* (Columbia Univ. Press, New York 1963)

17.2 E. Garfield: Citation indexes for science: A new dimension in documentation through association of ideas, Science **122**(3159), 108–111 (1955)

17.3 V. Bush: *Science, the Endless Frontier. A Report to the President on a Program of Postwar Scienctific Research* (National Science Foundation, Washington 1945)

17.4 H. Moed, W. Glänzel, U. Schmoch (Eds.): *Handbook of Quantitative Science and Technology Research* (Kluwer, Dordrecht 2004)

17.5 K. Debackere, W. Glänzel: Using a bibliometric approach to support research policy making: The case of the Flemish BOF-key, Scientometrics **59**(2), 253–276 (2004)

17.6 W. Glänzel, A. Schubert, B. Thijs, K. Debackere: A new generation of relational charts for comparative assessment of citation impact, Scientometrics **78**(1), 165–188 (2009)

17.7 W. Glänzel, K. Debackere: On the opportunities and limitations in using bibliometric indicators in a policy relevant context. In: *Bibliometric Analysis in Science and Research: Applications, Benefits and Limitations*, ed. by R. Ball (Forschungszentrum Jülich, Jülich 2003) pp. 225–236

17.8 A. Pritchard: Statistical bibliography or bibliometrics, J. Doc. **25**(4), 348–349 (1969)

17.9 V.V. Nalimov, B.M. Mulchenko: *Scientometrics* (Nauka, Moscow 1969)

17.10 G. Sivertsen: Publication-based funding: The Norwegian model. In: *Research Assessment in the Humanities Towards Criteria and Procedures*, ed. by M. Ochsner, S.E. Hug, H.D. Daniel (Springer, Zürich 2016) pp. 79–90

17.11 T. Braun, W. Glänzel, A. Schubert: *Scientometrics Indicators. A 32-Country Comparative Evaluation of Publishing Performance and Citation Impact* (World Scientific, Philadelphia 1985)

17.12 W. Glänzel, B. Thijs: The influence of author self-citations on bibliometric macro indicators, Scientometrics **59**(3), 281–310 (2004)

17.13 T. Braun, W. Glänzel: United Germany: The new scientific superpower?, Scientometrics **19**(5–6), 513–521 (1990)

17.14 H.F. Moed, R.E. de Bruin, N. van Leeuwen: New bibliometric tools for the assessment of national research performance: database description, overview of indicators and first applications, Scientometrics **33**, 381–422 (1995)

17.15 W. Glänzel: Characteristic scores and scales. A bibliometric analysis of subject characteristics based on long-term citation observation, J. Informetr. **1**(1), 92–102 (2007)

17.16 W. Glänzel, B. Thijs, K. Debackere: The application of citation-based performance classes to the disciplinary and multidisciplinary assessment in national comparison and institutional research assessment, Scientometrics **101**(2), 939–952 (2014)

17.17 W. Glänzel, A. Schubert: A new classification scheme of science fields and subfields designed for scientometric evaluation purposes, Scientometrics **56**(3), 357–367 (2003)

17.18 W. Glänzel, B. Thijs: Using 'core documents' for detecting and labelling new emerging topics, Scientometrics (2012), https://doi.org/10.1007/s11192-011-0591-7

17.19 L. Butler: What happens when funding is linked to publication counts? In: *Handbook of Quantitative Science and Technology Research*, ed. by H.F. Moed, W. Glänzel, U. Schmoch (Springer, Dordrecht 2004) pp. 389–405

17.20 J. Bichteler, E.A. Eaton: The combined use of bibliographic coupling and co-citation for document retrieval, J. Assoc. Inf. Sci. Technol. **31**(4), 278–282 (1980)

17.21 K.W. Boyack, R. Klavans: Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately?, J. Assoc. Inf. Sci. Technol. **61**(12), 2389–2404 (2010)

17.22 R.R. Braam, H.F. Moed, A.F.J. van Raan: Mapping of science by combined cocitation and word analysis, Part I: Structural aspects, J. Assoc. Inf. Sci. Technol. **42**(4), 233–251 (1991)

17.23 R.R. Braam, H.F. Moed, A.F.J. van Raan: Mapping of science by combined cocitation and word analysis, Part II: Dynamical aspects, J. Assoc. Inf. Sci. Technol. **42**(4), 252–266 (1991)

17.24 P. Glenisson, W. Glänzel, F. Janssens, B. de Moor: Combining full text and bibliometric information in mapping scientific disciplines, Inf. Process. Manag. **41**(6), 1548–1572 (2005)

17.25 K. Jonkers, T. Zacharewicz: *Research Performance Based Funding Systems: A Comparative Assessment*, JRC Science for Policy Report (European University Association, Brussels 2016)

17.26 E. Pruvot, A. Claeys-Kulik, T. Esterman: *Designing Strategies for Effective Funding of Universities in Europe*, EUA Report (European Union, Brussels 2015)

17.27 K. Debackere, M.A. Rappa: Institutional variations in problem choice and persistence among scientists, Res. Policy **23**(4), 425–441 (1994)

17.28 Y. Gingras: *Bibliometrics and Research Evaluation: Uses and Abuses* (MIT Press, Cambridge 2016) p. 136

17.29 W. Glänzel: Seven myths in bibliometrics. About facts and fiction in quantitative science studies, Collnet J. Scientometr. Inf. Manag. **2**(1), 9–17 (2008)

Koenraad Debackere

ECOOM and Faculty of Economics and Business
KU Leuven
Leuven, Belgium
koenraad.debackere@kuleuven.be

Koenraad Debackere is Professor of Innovation Management and Policy at KU Leuven. He is promotor-coordinator of the Centre for Research and Development Monitoring (ECOOM). His research has focused on the development and use of indicators measuring the effects of science and innovation policy, the role of entrepreneurial universities in economic development, and the management and growth of knowledge-intensive companies.

Wolfgang Glänzel

ECOOM and Faculty of Economics and Business
KU Leuven
Leuven, Belgium
wolfgang.glanzel@kuleuven.be

Wolfgang Glänzel is Director of the Centre for R&D Monitoring (ECOOM) and Full Professor at KU Leuven. He is also affiliated with the Department. Science Policy and Scientometrics at the Library of the Hungarian Academy of Sciences in Budapest (Hungary). Wolfgang Glänzel studied mathematics at Eötvös Lorand University (ELTE) in Budapest. He holds a Doctorate in Mathematics from ELTE, obtained in 1984, as well as a PhD in the Social Sciences obtained from Leiden University (The Netherlands) in 1997.
His research activities comprise probability theory and mathematical statistics, quantitative science studies, and research policy. Wolfgang Glänzel has published numerous journal articles, proceedings, and book chapters, and has co-authored/co-edited several books. He is Secretary-Treasurer of the *International Society for Scientometrics and Informetrics* (ISSI), *Editor-in-Chief of the international journal Scientometrics, and Academic Editor of the multidisciplinary OA journal PLoS ONE.*

Bart Thijs

ECOOM
KU Leuven
Leuven, Belgium
bart.thijs@kuleuven.be

Bart Thijs is Senior Researcher in bibliometrics at KU Leuven – ECOOM. His main research focuses on the identification of global and local network structures within scientific communication and on the development and application of citation-based indicators. In his work, he combines a background in mathematical psychology (master's degree at KU Leuven) and social sciences (PhD at Leiden University) with professional experience as a statistical consultant.

18. Different Processes, Similar Results? A Comparison of Performance Assessment in Three Countries

Sybille Hinze, Linda Butler, Paul Donner, Ian McAllister

Monitoring the scientific performance of a country, region, or organization has become a high priority for research managers and government agencies. Research assessments have been implemented to provide evidence and facilitate their decisions. They differ in the methodologies applied, the disciplinary and regional breadth, and the consequences that follow. We sought to examine the extent to which quantitative, indicator-based analysis can contribute to identifying and better understanding the effects and effectiveness of the different assessment regimes. To this end, we analyzed the publications from three countries (Australia, the United Kingdom, and Germany) with contrasting systems in place, seeking to demonstrate the possibilities and limitations of using an indicator-based methodology for determining the outcomes from different approaches to assessment.

We intentionally selected three countries with different assessment regimes, expecting to see the effects of this in the bibliometric analyses we undertook. However, we found that the data alone do not allow us to conclude that any one system has a beneficial or detrimental influence on performance. Rather, the data suggest that it is not the specific system that makes a difference but the fact that performance becomes a central topic of conversation.

In order to better understand the mechanisms behind changing performance, restricting scrutiny to mere numbers is insufficient. Contextual information at various levels of aggregation—within and outside the institutions—is highly relevant.

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_18

18.1 Background

With changed governance mechanisms in science, the evaluation of university research performance is now widely used and accepted as an essential management tool. Evaluation and performance monitoring systems have been developed and put in place in a number of countries. While these systems differ in the methodologies they employ to evaluate performance, bibliometric analyses are increasingly being used, either as a central building block or as a component to inform peer review. Although one of the most well-known performance-based funding systems, the Research Assessment Exercise (RAE) in the United Kingdom, only uses bibliometrics as a minor element to inform the peer review process, other countries or regions have developed systems to allocate research funding that are more explicitly based on these quantitative indicators, such as Norway, Denmark, Finland, and Flanders.

In Australia the Excellence in Research for Australia (ERA) framework is based on reviews by experts where, for science disciplines, information is provided in the form of a range of indicators, including bibliometric measures, and no additional peer review of individual outputs is undertaken. However, while university research performance is assessed by ERA, little funding is reallocated as a consequence of the outcomes of these reviews.

While both the British RAE and the Australian ERA systems use subject-based panels of experts, they differ significantly with regard to what publications are assessed. In the United Kingdom only the self-selected *best four* publications are submitted, while in Australia the total academic output is taken into account.

In Germany, performance assessment systems for universities, using different methodologies, have been implemented at Länder level rather than the federal level. An exception is the Research Rating developed and implemented by the German Council of Science and Humanities (Wissenschaftsrat), though only a very limited number of exercises have been carried out. The Research Rating has not progressed beyond the pilot exercise phase and it has yet to be employed on a systematic basis. However, research-performing sectors in Germany are monitored at the federal level via the Pact for Research and Innovation (PFI). Bibliometric performance indicators are regularly constructed and form one part of the overall monitoring of research institutions. In addition, the Research Core Data Set (RCD) was recently developed with the intention of enabling decentralized data collection, including publication data. The RCD, despite the fact that its implementation is voluntary, could enable monitoring research performance at institutional level.

The highly aggregated PFI bibliometric analysis in Germany has, to this point, relied on extracting institutional data directly from one of the well-known commercial suppliers. In contrast, the exercises directed at the institutional level in the United Kingdom and Australia, as well as Länder-level exercises in Germany, draw on data submitted by the institutions themselves.

In this chapter we explore, compare, and assess the outcomes of different approaches. At the national level, we briefly examine two countries where performance monitoring systems are already in operation, namely the British RAE and the Australian ERA; we also examine Germany, where such a system was pilot-tested but has not yet been implemented on a national basis.

Using these countries for comparison enables us to assess outcomes based on three different systems:

- One where data collection is highly selective and is linked to funding (United Kingdom)
- One based on comprehensive data sets provided by institutions but not linked to funding (Australia)
- One without a broad national assessment system where, when used, publication data is primarily extracted centrally from the commercially available data bases (Germany).

We also look at a selective exercise at the Länder-level in Germany, where data sets are provided by institutions, but the outcomes are not linked to funding.

We examine the operation of these three national systems at the institutional level, using chemistry as a case study. For this discipline, we have information available in Germany from a Länder-level assessment in Lower Saxony, and for all three countries we can select institutions that have been the subject of performance assessment.

In the next section, we provide a brief introduction to the systems that are in place to monitor research performance in the three countries. As much has been written on the United Kingdom RAE, as well as on the Australian ERA, we will provide more detail on the exercises implemented in Germany, as these are less well known internationally.

18.2 Research Assessment in the United Kingdom

18.2.1 A Brief History

The United Kingdom has a long history of assessing the research performance of its universities, with the first steps towards establishing a national assessment system dating back to 1985. The United Kingdom University Grants Committee (UGC) announced that the first explicit and formalized assessment process of the quality of research would take place in 1986. Nearly a decade followed before other countries followed the United Kingdom's lead.

In the first *research selectivity exercise*, universities submitted five outputs (publications and/or patents) in 37 cost centers, along with a four-page context statement highlighting the unit's strengths. The assessments were undertaken by the UGC's subject sub-committees. A second selectivity exercise was undertaken in 1992, with universities submitting up to two publications for every member of staff covering 152 subject areas. These were assessed by 70 peer review panels, which were external to the Universities Funding Council, which had taken over the role from the now defunct UGC.

The United Kingdom's national assessment exercise was first known as the Research Assessment Exercise (RAE) in 1992; in this round, universities were able to submit up to four outputs per *research active* staff member, and a rating scale was used to evaluate units. Variations to the rating scale were introduced in the next two exercises (1996, 2001).

Concerns over the large cost of the RAE were raised, particularly by the United Kingdom Treasury Department. As a result, after the 2001 exercise, there were moves to have it replaced by a system that relied more heavily on metrics such as citations, research income, and postgraduate numbers, rather than peer review. From the beginning, there was scepticism among relevant actors in the system, including the Higher Education Funding Council for England (HEFCE) and individual institutions, which led to the rejection of a metrics-based system. This scepticism remains, as detailed in the recently carried out review of the *Role of Metrics in Research Assessment* [18.1]. Another concern of the existing methodology was that large universities could hide a long tail of lesser work and still obtain a 5* (top) rating and have all of their academics funded at this level. The 2008 RAE saw the introduction of quality profiles for each unit of assessment, rather than a single grade.

The last RAE was conducted in 2008. The national assessment exercise was re-branded, becoming the Research Excellence Framework (REF), with the first of the new assessments undertaken in 2014. It was broadened to include an assessment of the impact of research, and the environment in which it was undertaken, in addition to the quality of research outputs.

18.2.2 The Assessment Process

Since 1996, the research output of universities has been assessed by between 36 and 69 panels of experts. Institutions submitted up to four outputs per research active staff member across a 6 year period. *Outputs* included many forms of publication, such as journal articles, monographs, and chapters in books, as well as outputs disseminated in other ways such as patents, designs, performances, and exhibitions.

While the unit of assessment generally aligns with a department, school, or faculty within a university, it is not always a precise alignment—submitted staff (and, therefore, their outputs) may come from a range of organizational units drawn from across the organization. It is the university's choice to determine the panel that best fits their staff members' research.

The panels then assessed the quality of outputs against the criteria of originality, significance, and rigor. For all exercises, the assessment was based on peer review of the submitted outputs. In the 2014 REF, some panels were provided with basic citation information on the number of times a journal article was cited, as well as contextual information to support peer review.

Until the 2008 RAE, units of assessment were given a single star rating for the overall quality of their submission. The rating scale varied over time—four points (1986), five points (1992, 2008), and seven points (1996, 2001). In RAE 2008, quality profiles were introduced and panels determined the proportion of a unit's submission that fell into each of the five categories. Quality profiles across a five-point scale were retained for REF 2014 [18.2].

While the assessment of research output remained the central element of the process, from 2014 universities were also assessed on the impact their work had on *the economy, society, culture, public policy or services, health, the environment, or quality of life, beyond academia*. They were also assessed on the research environment, that is, the strategy, resources, and infrastructure that support research [18.3]. Each element contributing to the overall assessment would attract a discrete proportion of the funding—65% for outputs, 20% for impact, and 15% for the environment.

18.3 Research Assessment in Australia

18.3.1 A Brief History

There was a major overhaul of the Australian university sector in 1988 when the binary divide between universities, on the one hand, and colleges of advanced education and institutes of technology, on the other, was removed. Many institutions were merged, and all the organizations that were created by the reform were classified as universities. The original research quantum (RQ) was introduced in 1990 in order "... to provide a more equitable funding" to universities in the new national system [18.4]. However, from its inception, changes to the RQ were mooted as it was based solely on data for Commonwealth Competitive Research Grants, which favored the pre-reform universities; indeed, from 1990 to 1994, RQ allocations continued to go only to those universities.

The broadening of the index that underpinned the RQ was the subject of extensive consultations. In 1991, the government commissioned Russell Linke to undertake a review of performance indicators in higher education, in part to feed into changes that were anticipated to the RQ [18.5]. In late 1993, a working party was set up to develop a new index [18.4]. This new index consisted of three components—success in competitive grants, graduate student numbers, and publication output.

Australian higher education funding and assessment policies continued to be modified throughout the 1990s and into the first decade of the new millennium. The *Kemp* review at the end of the 1990s made it clear that while the RQ might, for better or worse, be acceptable for its purpose—the distribution of the block research grant among universities—but that it was a very poor instrument for assessing the performance of universities [18.6]. A close watch was being kept on developments in the United Kingdom, and Sir Gareth Roberts was asked to chair a committee to draw up a framework for assessing universities' research performance drawing on his review of the United Kingdom's research assessment exercise. The government's attention turned to research assessment, as distinct from mere funding distribution. The Research Quality Framework (RQF), later to be superseded by the ERA initiative, was developed. ERA is now a well-entrenched national research assessment exercise. The first pilot test of ERA, overseen by the Australian Research Council (ARC), was undertaken in 2009, followed by full-scale exercises in 2010, 2012, 2015, and 2018.

While ERA has little direct impact on government funding allocations (though its predecessor the RQF did intend to use it to distribute RQ money), Australian universities are heavily reliant on international students to generate income [18.7]. These students take considerable notice of both the international rankings and the detailed data published on ERA outcomes, so while they have no direct government funding implications, they do have significant financial implications.

18.3.2 The Assessment Process

The ERA assessment process was developed after extensive consultations between the ARC and the higher education sector. This involved setting up working groups in a range of disciplines and establishing an over-arching metrics working group drawing on local and international experts. Minor refinements to various aspects of the submission and assessment process have been made, but the suite of indicators prescribed in the discipline matrix of the framework remains largely unchanged.

Evaluation of data submitted by the universities for ERA 2015 was undertaken by eight Research Evaluation Committees (RECs), broadly representative of eight discipline clusters. The ERA 2015 RECs were comprised of 155 distinguished researchers from Australia and overseas, with expertise in their fields and in research evaluation. Their task was to assess the overall performance of universities for fields and sub-fields of research in which they were active. The decision to focus the ERA assessments on fields rather than departments or groups was taken after extensive consultations within the higher education sector and a detailed examination of the strengths and weaknesses of other systems, particularly the United Kingdom's RAE. ERA is based on the principle of expert review informed by indicators. Evaluations were performed by four broad categories of indicators covering quality, activity, application, and recognition.

Research quality was considered on the basis of a publishing profile, citation analysis, ERA peer review, and peer-reviewed Australian and international research income. ERA did not attempt to develop a uniform method of assessment, but rather developed a suite of indicators that varied across disciplines (known as the *discipline matrix*) [18.8]. For example, citation indicators were only used for relevant fields (i. e., those where at least 50% of the total output in the discipline was indexed by the selected citation data supplier), predominantly the sciences, while peer review of submitted outputs was undertaken for most disciplines in the social sciences, the humanities, and the arts.

Universities reported all the outputs of staff with appointments classified as research only or research and teaching. Where used, citation analysis was undertaken on the university's full publication set in the relevant field. Where peer review was used, the assessment was undertaken on 30% of the submitted publications from the university, which the institution itself nominated [18.9].

Universities were given an overall grade between 1 (low) and 5 (high) for each field and sub-field in which they were active, but individual ratings were not given for the four separate elements of the assessment—quality, activity, application, and recognition.

18.4 Research Assessment in Germany

18.4.1 National Level

The organization of research assessment in Germany is largely decentralized. It most commonly takes place either at the regional (Länder) level or at the institutional level, and hence comparative analyses are scarce. A few exceptions exist, such as the Research Rating, which was developed by the German Council of Science and Humanities (*Wissenschaftsrat*—WR), which undertook a critical analysis of national and international assessment methodologies [18.10]. While assessment of comparative research performance was seen as an "essential component of a reform process that reinforces the autonomy of academic institutions and involves a transformation from detailed state control to a global system comprising elements of competition" [18.10, p. 34], introducing a ranking based solely on quantitative data and indicators was not recommended in Germany.

Instead of quantitative indicators, an approach based on informed peer review that incorporated qualitative, quantitative, and contextual information was recommended. The process used by the Research Rating consisted of setting up a steering group and establishing assessment panels whose task it was to assess the performance of the institutions under review in the relevant field. It also outlined the assessment dimensions and criteria: research quality, impact and efficiency; promotion of young researchers—processes and success; and knowledge transfer—relevance, application in businesses, further education, and research-based consultancy. In order to take into account disciplinary differences, each assessment panel detailed the criteria to be applied and, in consultation with the steering group, defined the rating scale. This also included the decision as to which data, qualitative and quantitative, were to be applied to each criteria. It also included the decision on whether or not to consider bibliometric data. The exercise yielded a report reflecting a performance profile across the assessed dimensions for each institution.

The main objective of the Research Rating was not to distribute funds but to

> provide support for universities and non-university research institutions both in their missions and—in connection with other quality assurance and strategic planning procedures—in quality assurance measures in research, and to promote competition for quality. [18.10, p. 44]

Since its introduction, pilot studies have been carried out for chemistry (2007), sociology (2008), electrical engineering (2011), and English and American studies (2012), but the approach has not been introduced systemically. Besides criticism regarding continual evaluation and its unintended effects, the general appropriateness of the approach, the information and data used, data availability, and the efforts needed to make them available were all the subject of criticism.

Experiences gained from the pilot exercises were taken into consideration in the Council's recommendations about the future of the Research Rating [18.11] and consequently the recommendations towards developing a Research Core Data Set (RCD) [18.12]. The RCD aims to standardize data formats, making data sharing between all actors a much easier and more straightforward task. It is targeted at both research-performing institutions and report-requesting organizations and, by standardizing the reporting processes, should reduce the burden when it comes to responding to various internal and external information requests and, at the same time, ensure comparability of data.

The first version of the specification for the RCD was published in January 2016 [18.13] following discussions about its content (the data and research information to include) and its operationalization (the definition of the concepts). The RCD is not an actual data set as such, and no centralized data collection will be initiated. Rather, the RCD sets out rules and recommendations for data collection and processing in the research-performing institutions themselves. It covers input, output, and process data such as human resources, research projects and third-party funding, patents, spin-offs, and research infrastructures, as well as publications [18.14].

Universities and research organizations can decide upon the implementation of the RCD on a voluntary basis. As the RCD research is still in its infancy, it is too early to comment on the extent of its uptake, its use in the context of research assessments at various levels, or its potential effects.

18.4.2 Länder Level

Performance assessment at the Länder level varies widely. By 1999 the *Wissenschaftliche Kommission Niedersachsen* (WKN, English: Academic Advisory Council Lower Saxony) had implemented a framework to systematically assess research performance at universities and in research organizations [18.15]. At the time, and still to this day, it remains a unique exercise at the Länder level.

The WKN is an independent expert committee which, based on its evaluations, provides advice to the state government as well as the state-financed universities and research organizations in regard to further development, monitoring, and assessment of the state's higher education and research system, including its structure and research profile [18.16]. Initially, between 2000 and 2008, all disciplines represented at universities in Lower Saxony were successively evaluated. Approximately 3 to 4 years after the initial evaluation, an assessment of the implementation of the evaluation recommendations was undertaken (for an overview of the evaluations undertaken see [18.17]). From 2015 onwards, the system has used a demand-driven approach rather than evaluating all disciplines regularly. In order to ensure high quality research, a monitoring system aimed at identifying relevant disciplines for which an evaluation should be initiated was introduced. In addition, complementing the disciplinary approach, topic-focused evaluations were introduced (for an overview, see [18.18]). In addition to these initiatives, the WKN also carries out institutional evaluations.

Evaluations are carried out as *informed peer review* by an independent expert group. Universities active in the discipline or topic under evaluation are included. Self-reports are prepared by the universities based on the guidelines provided, and there are site-visits by the expert group. The evaluation concept prescribes quality and relevance of research and effectiveness and efficiency as the two criteria to be met. In their self-reports, universities also submit data on research output, third party funding, early career researchers, international visibility and networking, and prizes and awards.

To ensure discipline-relevant evaluations, each expert group defines a set of discipline-specific criteria and their operationalization. Bibliometric data and analysis can be taken into account and were utilized in the recent evaluation in chemistry, which was carried out in 2015. The assessment provided by the expert group is the basis for the final evaluation report, which is written by the WKN office and agreed upon with the expert group. As at the federal level, the evaluation has no direct consequences for funding, but instead provides recommendations and initiates processes of organizational learning and change.

18.4.3 Pact for Research and Innovation

While bibliometric data plays only a marginal role in the Research Rating, it is used systematically in monitoring the PFI. The PFI, initiated in 2005, is one of the major instruments designed to strengthen the performance of the German science system. Providing additional funds for the four large non-university research organizations (Max Planck Society, Fraunhofer Society, Helmholtz Association, and Leibniz Association) as well as the German Research Foundation (DFG), it complements the Excellence Initiative and the Higher Education Pact, both of which are primarily focused on higher education institutions. A monitoring report is published annually, presenting a wide range of indicators, including an extensive bibliometric analysis.

The bibliometric analyses for the PFI are carried out centrally and are made publicly available via the website of the German Federal Ministry of Education and Research (for the 2016 report, see [18.19]). The report provides comparative information with regard to the performance of Germany as well as for the large research organizations and the universities. The analysis of performance is presented at an aggregate level. For universities, this means that the whole university sector is analyzed, rather than individual institutions, as the aim is to avoid constructing a league table. The same applies to the non-university organizations.

While the monitoring, which is carried out regularly, does not have direct and immediate consequences on the funding of the organizations, it is a way of reviewing whether the additional funding provided via the PFI yields any effects on the sector level.

18.5 Comparing What is Assessed in Each System

18.5.1 The United Kingdom

In the 2014 REF, and in the RAEs that preceded it, universities submitted units of assessment (UoA), which may, but need not, comprise staff who work within a single department or other organizational unit at the university. A submitted unit may comprise staff who work in multiple organizational units at the university [18.20]. It is, therefore, uncertain exactly what the REF assesses. It is not clear whether it is an organizational unit as defined within the university's structure, or all the output in a field as defined by the REF panel structure. Which of the two scenarios comes closest to describing an institution's submission will vary from university to university.

Universities submit only their *best four* outputs from the 6-year assessment period for each current research active member of staff, not their total output, and also select to which panel an academic's work is submitted. The effect of these two choices, looking specifically at our case study of chemistry, is:

1. Not all output in the chemical sciences is submitted for assessment.
2. Not all output in the chemical sciences is submitted to the chemical sciences panel (panel 8).
3. The output submitted to a panel does not necessarily (though it may) represent one academic department, school, or faculty.

In addition, the outputs submitted need not necessarily have been published while the staff member was employed by the university.

Given the above, to then interpret the results from panel 8 as showing the strength of chemical sciences, or schools of chemistry, in the different universities is problematic.

18.5.2 Australia

In the ERA exercise, the UoAs are fields of research as defined by the Australian and New Zealand Fields of Research Classification Scheme (FoR) [18.21]. Evaluation occurs at both the two-digit field (i. e., such as chemistry, physics, etc.) and the four-digit sub-field (i. e., such as organic chemistry, nuclear physics, etc.) FoR level for each institution that is considered research active. Universities must submit all publications for assessment for each FoR they are deemed active under ERA rules.

The ERA journal list forms an integral part of the ERA evaluation process. The list is used to define, with limited exceptions, the eligible FoR codes that research outputs may be assigned to during the submission phase. The journal lists also form the basis for the calculation of international and Australian benchmark figures.

In contrast to the United Kingdom REF exercise, Australian universities are assessed on all publications and have little freedom to choose the panel that assesses them. However, in one respect ERA and REF are the same—publications are not restricted to those based on research carried out at the university but can include publications authored before a staff member joined the institution.

18.5.3 Germany

The Research Rating of the WR is not comprehensive in the sense that all disciplines or fields of science are examined in a single exercise. By contrast, PFI monitoring covers all discipline areas but is a highly aggregated exercise focusing on the development of the organizations as a whole and does not cover the disciplinary level. Thus, while this exercise needs to be taken into account when discussing national research performance assessment systems, it does not provide assessment outcomes that allow us to make institutional comparisons with universities from Australia and the United Kingdom. For this reason, the discussion that follows will focus on the regional exercise implemented in Lower Saxony.

The evaluation in Lower Saxony is discipline focused and entails delimiting the field based on the organizational units active in the discipline. Like in the United Kingdom and Australia, these might also be located in different faculties. For the bibliometric analysis, which was carried out as part of the evaluation of chemistry, approximately 110 research groups in 24 institutes in 6 universities in Lower Saxony participated in the assessment. The overall assessment was carried out by an expert panel consisting of six members.

As has already been mentioned, each university had to provide a self-report including data on research output, third party funding, early career researchers, international visibility and networking, and prizes and awards. Information on research output included submitting publication lists per research group. In addition, the bibliometric analysis, which was carried out centrally, provided bibliometric indicators for each of the groups (Sect. 18.3.2).

18.6 Comparing the Role of Metrics in Each System

18.6.1 United Kingdom

While there has been much discussion over many years in the United Kingdom on the introduction of metrics into the assessment process, from Gareth Robert's review in 2003 to HEFCE's consultations in response to treasury requests from 2006 to 2009, and to the *Metrics Tide Report* in 2015 [18.1], ultimately they have had little impact on the process.

Citation data was introduced into panel deliberations for some of the sciences in the 2014 REF, but the only data provided were simple counts of citations for each publication. The way in which panels could use the data was very limited [18.22]:

> 62a. Where available and appropriate, citation data will form part of the process of assessment, in relation to the academic significance of outputs. It will be used as one element to inform peer-review judgements made about output quality, and will not be used as a primary tool in the assessment.

However, it was acknowledged in the Wilsdon et al. report that citation data did enter in the evaluation process where marginal judgments were concerned [18.20].

After the release of the *Metrics Tide Report* in 2015, it is unlikely that metrics will play any greater role in the 2018 exercise. There remains a strong belief in the United Kingdom that citation analysis (the main suite of metrics on which consultations took place) cannot be uniformly applied and is, therefore, inappropriate. The United Kingdom does not appear ready or willing to go down the path taken by Australia of developing a discipline-specific set of indicators.

18.6.2 Australia

In ERA, quantitative indicators played a pivotal role in the assessment of research in those fields where citation data were used. The information provided went well beyond the simple counts used in REF 2014 [18.9, 23]. Using data from Elsevier's Scopus database, three bibliometric profiles were provided to assessment panels:

1. Relative citation impact (RCI), calculated against Australian university and world benchmarks
2. Distribution of papers based on world centile thresholds and Australian university averages (centile analysis)
3. Distribution of papers against RCI classes.

The three indicators were designed to be considered as a complementary set and not used individually. Where citation data were deemed appropriate, no peer review of individual outputs was undertaken.

The disciplines that relied on citation analysis rather than peer review were: physical, chemical, mathematical (excluding pure mathematics), earth, biological, medical and health sciences, nanotechnology and biotechnology, and engineering.

18.6.3 Germany—The Example Lower Saxony

The chemistry evaluation referred to in this paper included an extensive bibliometric analysis. The universities were asked to submit publication lists for all research groups, which were subject to the evaluation. Clear guidelines were given as the data provided was used for matching and retrieving the publications from the Web of Science.

For the analysis, the in-house data infrastructure set up and run by the German Competence Center for Bibliometrics was used, with the Web of Science as the data source. Submitted publications were identified based on their DOI (digital object identifier) or, where no DOI was available, using a match key consisting of publication year, title, name of the first author, and source title. Around three-quarters of the publications submitted were identified by the matching procedure applied. The majority of non-identified publications were book chapters, proceedings papers, and German language publications. For the bibliometric analysis, only articles published in the period 2008–2013 were taken into consideration, and only if at least one author was a member of the unit of analysis within this period.

To inform the assessment of the performance of the institutions, a range of indicators were used:

- Publication data (number of publications, number of publications per research staff)
- Citation based impact indicators (average citation rate, field normalized citation rate, share of highly cited publications, and share of uncited papers)
- Co-authorship based indicators to assess extent and patterns of collaboration.

Benchmarks were provided for comparison with institutions in Lower Saxony and for Germany in total.

While the bibliometric data was taken into consideration to inform the panel, its representativeness was questioned by the expert panel due to the incomplete coverage. Overall, bibliometric data did not play a central role in the assessment, rather the assessment was reached based on bringing together qualitative information gathered from the institutional self-reports, the site-visits and data reflecting various performance dimensions as well as context information.

In summary (see Table 18.1), the three evaluation systems examined here show considerable differences with regard to the functions they fulfil and the methodologies applied. The question that arises from this is whether they also lead to different outcomes in terms of performance. Do we see differences and thus varying effects, which might be explained by the different approaches? Or, do different approaches to performance assessment result in similar outcomes? To address these questions we carried out a bibliometric analysis to find evidence that would allow us to draw conclusions on the effectiveness of the systems. Our primary focus is the institutional level, but we also present contextual data at the national level. We employ basic and widely used bibliometric indicators.

Table 18.1 Comparison of the assessment regimes of Australia (ERA), Germany (WKN), and the United Kingdom (REF)

	Australia	**Germany**	**United Kingdom**
Assessment system	ERA—National	WKN—regional (Länder) level-single discipline	REF
Format of assessment	For the sciences—metrics based assessment by expert panels; other disciplines—peer review by expert panels, also informed by a range of discipline-specific indicators	Assessment by expert panels informed by a range of discipline-specific indicators	Peer review with some contextual data
Role of bibliometrics	In the sciences, the central information used in the assessment of quality was drawn from a range of citation measures, supplemented by other indicators as set out in ERA's discipline-specific indicator matrix	For chemistry, extensive bibliometric data was provided including a range of citation measures and co-authorship information, but limited use of the data in the overall exercise	For the sciences, citation counts for each submitted journal article were provided. However, they were only to be regarded as contextual information. Some panels appear to have used them to resolve disputed assessments
Years conducted	2010, 2012, 2015. Next one is scheduled for 2018	Chemistry: 2000, 2005, 2015 (bibliometrics only used in the latest exercise)	1986, 1989, 1992, 1996, 2001, 2008, 2014. Next one scheduled for 2018
Funding implications	Only linked to a very small amount of funding	No direct link to funding decisions	65% of funding attached to the REF was tied to the assessment of outputs
Rating scale	5 well above world standard 4 above world standard 3 at world standard 2 below world standard 1 well below world standard n/a due to low volume	Descriptive rather than fixed scale	4* world leading 3* internationally excellent 2* recognized internationally 1* recognized nationally unclassified—below 1* standard or not classified as research
Unit of assessment	All outputs classified to a field. For chemistry, relevant outputs were determined by the journal in which they appeared. Universities had some discretion when it came to articles in multidisciplinary journals	All university based research groups active in the relevant discipline	Universities have freedom to decide which panel to submit researchers to—some submit whole departments/schools to one panel; others split staff across two or more panels
Publications assessed	Universities must submit all outputs from every research active staff member	All the publications of relevant research groups are submitted to the assessment panel	The *best four* publications for each research active staff member submitted
Institutional affiliation of work submitted	Publications of all current staff, irrespective of where research undertaken	Publications of all staff during their employment in units within the period of assessment submitted	Publications of all current staff, irrespective of where research undertaken
Identification of publications	Universities submit publication data	Universities submit publication data	Universities submit publication data

18.7 Data and Methods

For the analysis we used the in-house data infrastructure of the German Competence Centre for Bibliometrics, a consortium consisting of seven institutions funded by the German Federal Ministry of Education and Research. The data infrastructure consists of the Web of Science as well as Scopus and is hosted at the Leibniz Institute for Information Structure Karlsruhe (FIZ Karlsruhe). Our analysis used Scopus data and analyzed a subset that was restricted to:

- Publication years 1996 to 2013
- Article and review document types
- Publications from the United Kingdom, Australia, and Germany
- Publications appearing in a pre-defined chemistry journal set.

The chemistry journal set was constructed by combining chemistry journals from the 2015 ERA journal list and journals classified by Elsevier to either a chemistry discipline or the broad field of chemistry. We did not include journals classified to two or more broad fields of research by Scopus.

Within the data set we also identified all publications for nine universities, three from each country, for which to undertake lower level analysis. These were:

- *Australia*: The Australian National University (ERA rating 4), the University of Melbourne, (ERA rating 5), and Monash University (ERA rating 5)
- *United Kingdom*: The University of Birmingham (REF 4*—13.8%, 3*—81.6%), the University of Bristol (REF 4*—28%, 3*—69.5%), and the University of Edinburgh (REF 4*—17.4%, 3*—74.1%)
- *Germany*: The Technical University of Braunschweig, the University of Hannover, and the University of Göttingen; all three were rated *excellent* by the expert panel [18.24].

The project data set contains 4 065 823 articles.

The bibliometric indicators constructed for the analysis were:

- Number of publications, limited to articles and reviews, whole count.
- RCI based on 3 year windows, with calculations limited to articles and reviews and specific to year of publication and world averages calculated.
- Identification of highly cited publications (top 1% and top 10%) and those that received citations above the world median—again calculated by year on 3-year citation windows and limited to articles and reviews.

Methodological note: For the full set of research articles and reviews based on the project's journal list, citation count threshold values for the calculation of shares of highly cited papers were computed. Citations were counted for the year of publication and the two following years, forming sliding 3-year citation windows. The 99th (90th, 50th) percentile value of the citation count distribution of each publication year was used as the threshold for definition of the class of the 1% (10%, 50%) most highly cited chemistry papers of that year.

In contrast to the Australian REF and in the Lower Saxony research evaluations, where information on exhaustive publication lists were taken into account, the United Kingdom REF assessment used a selective set. As we were able to extract a full output in the chemical sciences journals for our three target universities, it is informative to compare the results obtained for each university as a whole, with those obtained using the smaller sub-set of publications submitted to REF 2014. How much of the United Kingdom's universities' output in these journals was actually assessed in the REF?

Data for REF 2014 publications were downloaded from the HEFCE website [18.25]. Universities submitted 4699 outputs to REF panel 8 (chemical sciences) for assessment, of which 4689 were journal articles, one was a chapter in book, two were conference publications, three were patents, and four were software.

The fields to which the journals that the articles appeared in could be broadly classified were:

Chemistry	3654 (78% of all journal articles)
Multidisciplinary	616
Other Natural Sciences	318
Biomedical Sciences	77
Unclassified	33.

Using DOIs, 3138 of chemistry journal articles matched articles in our project data set i. e., 85.9% of articles in chemistry journals, and 67% of all journal articles. There are a number of reasons for the missing chemistry articles—universities were able to submit the output of research staff published prior to their employment at the university, so many of the missing publications did not contain a United Kingdom address. Also, a number of submitted outputs did not fall into our two publication categories—articles and reviews—and so were excluded from our project data set.

18.8 Analysis of Bibliometric Data

As all three countries have different assessment regimes in place, we might expect to see the effects of this in the bibliometric analyses we undertook. With a nationwide system linked to significant levels of funding, the United Kingdom might be expected to have the strongest performance. Then Australia, with a system-wide assessment exercise, though in this case not linked to funding, might show a performance somewhat less than the United Kingdom, but ahead of Germany where no broad assessment regime currently exists.

We used the data we had extracted from Scopus to investigate whether this hypothesis could be supported. First, we estimated overall trends for total publication output in each country, then examined lower levels of aggregations using chemical sciences as a test case. We chose chemical sciences because this was a field in which we had data at an even lower level of aggregation—the university—on which to perform detailed analyses for all three countries.

18.8.1 National Level—All Fields

The anticipated diverging trends in the national level data did not emerge; rather, we observed similar trends of moderately increasing publication numbers for Germany and the United Kingdom, while for Australia the increase is much more pronounced. Figure 18.1 shows these trends, using a logarithmic scale to highlight the differences between Germany and the United Kingdom, on the one hand, and Australia, on the other.

This result leads us to conclude that at the national level, there is no unambiguous effect on simple publication counts from the different evaluation systems. Attaching resource allocations to the exercise (United Kingdom only) does not lead to significant increases in research output compared to those systems where funding is not tied to the assessment exercise (Australia) or where there is no country-wide assessment regime (Germany).

As expected, when we move from publication counts to examine trends in the countries' shares of world output (Fig. 18.2) the results are analogous, with similar trends for the United Kingdom and Germany (decreasing shares), but the reverse trend for Australia (increasing share). The decreasing share of publications from Germany and the United Kingdom is not surprising, as it may well reflect a displacement effect as new entrants such as China and India emerge. What then is driving Australia's increased share?

To further investigate, we used the field of chemistry to try and determine whether the findings were replicated at lower levels of aggregation—at the level of field in a country and for individual institutions. We also wished to examine whether the introduction of citation analyses could highlight any trends that are not apparent from a simple count of publication numbers.

18.8.2 National Level—Chemical Sciences

In chemistry, the trends mirror those found for the overall national publication output for the three countries: very moderate increases for Germany and the United Kingdom with at the same time decreasing publication shares, while for Australia we find a more pronounced increase and increasing publication share (Figs. 18.3 and 18.4).

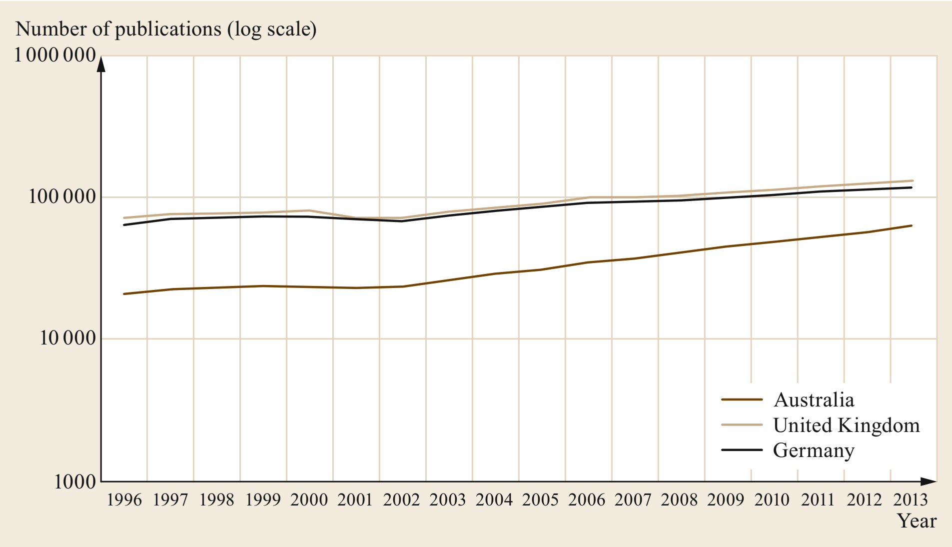


Fig. 18.1 Number of publications in all fields for the selected countries, 1996–2013. Data source: Scopus, authors' calculations

Part C | 18.8

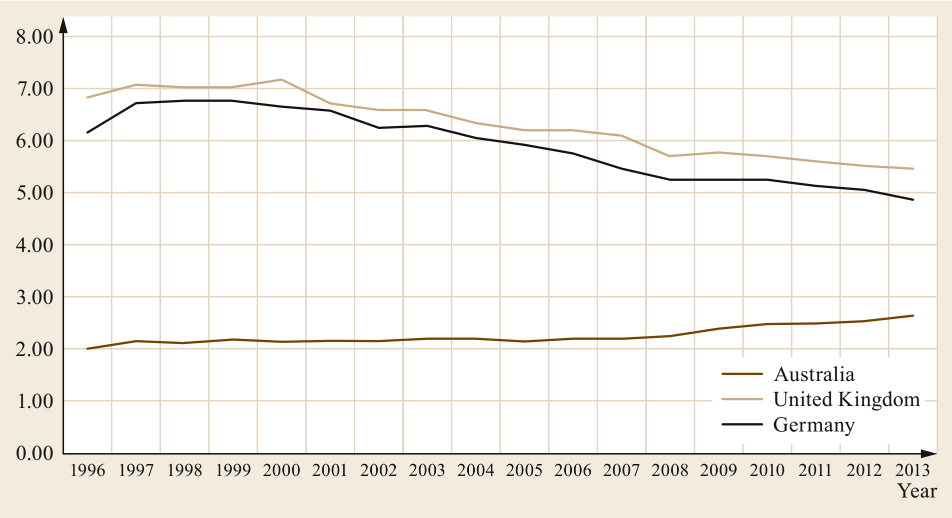


Fig. 18.2 Share of world publications in all fields for the selected countries, 1996–2013. Data source: Scopus, authors' calculations

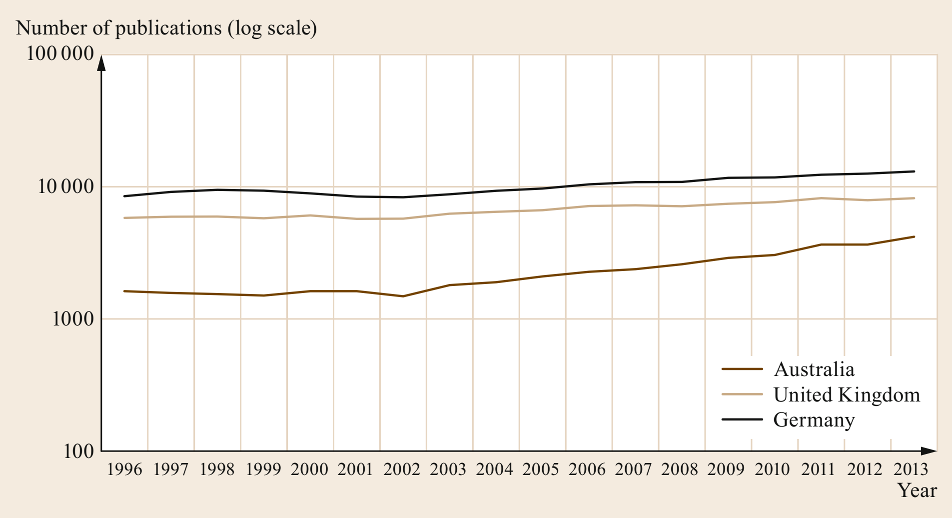


Fig. 18.3 Trends in number of publications in chemical sciences for the selected countries, 1996–2013. Data source: Scopus, authors' calculations

We examined trends in research performance for the three countries using two common citation measures—RCI (Fig. 18.5) and the proportion of output in the world's most highly cited publications (Fig. 18.6).

As with simple publication counts, the citation analysis highlights some counter-intuitive patterns, though all three countries remain at levels well above the world average. After some early volatility, the United Kingdom's RCI increased between 2002 and 2007 and has remained steady since. Allowing for yearly variances, Germany's RCI has remained unchanged across the 18 years of the analysis. In contrast, Australia's RCI increased significantly after 2002 and is now at a similar level to Germany, though at the start of the period it was well below Germany. Similarly, with their share of world publications in the top 10% most cited articles in the chemical sciences journals, Germany and the United Kingdom have been steadily declining since around 2000, while Australia's share has been rising. For all countries, the share of highly cited publications is above the share of total publications. This represents a strong performance.

Interestingly, the RCI for both the United Kingdom and Australia starts improving at the same time that policy discussions on the inclusion of citation measures in research assessment exercises begin to have currency. However, more analysis would need to be undertaken before any causal link between changes in policy and performance can be made, and the trends in shares of the top 10% of articles does not show the same consistency between countries.

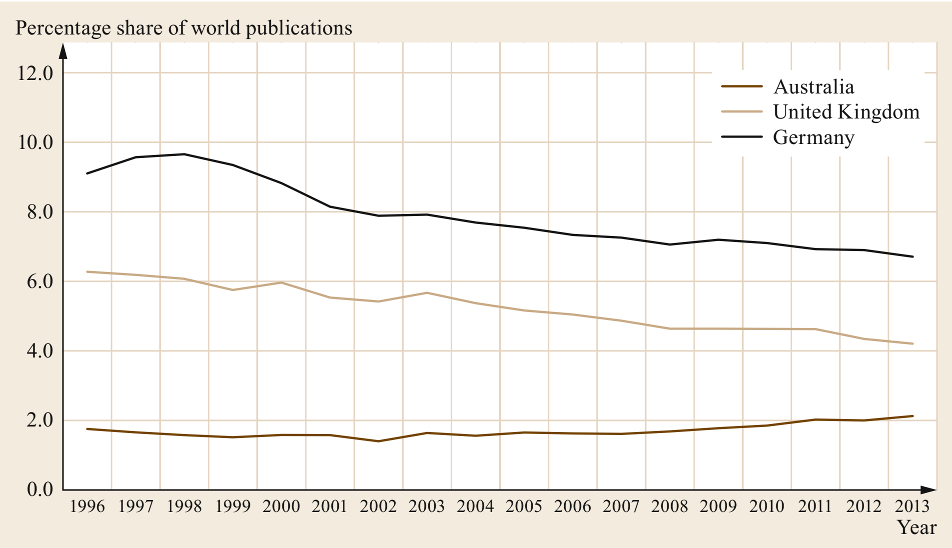


Fig. 18.4 Trends in share of world output in chemical sciences for the selected countries, 1996–2013. Data source: Scopus, authors' calculations

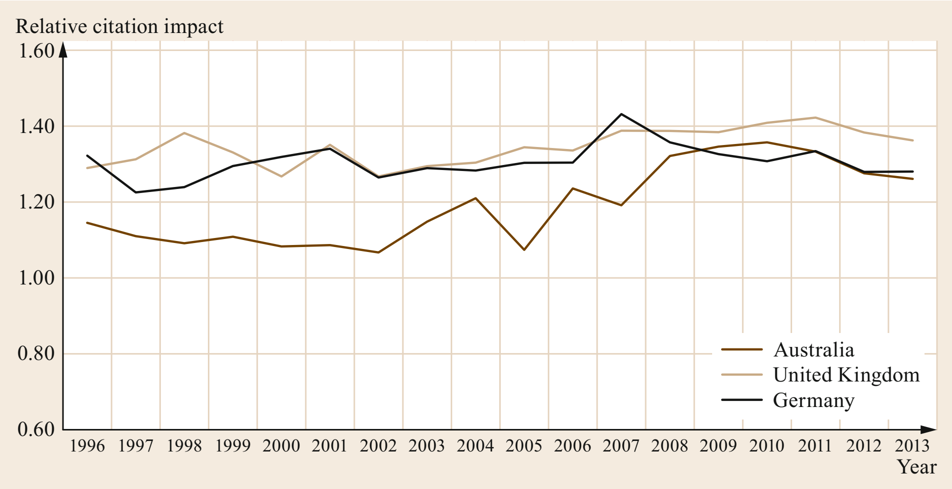


Fig. 18.5 Trends in relative citation impact for the selected countries in chemical sciences, 1996–2013. Data source: Scopus, authors' calculations

18.8.3 Institution Level—Chemical Sciences

While the overall developments found at the national level are mirrored by the selected institutions, varied performance among universities is also evident. As annual data at this lower level of aggregation is very volatile, the trends were analyzed using 3-year rolling windows.

For Germany, all three universities' publication numbers are increasing (Fig. 18.7), though for Braunschweig and Hannover, this was after a brief period of decline at the end of the 1990s. These three universities also increased the impact of their publications and, without assuming a strictly causal relationship, this improved performance occurred at the time of the initial evaluation in 2000 and in the period moving towards the follow-up exercise in 2005.

The steep increase in RCI for Göttingen at the end of the time frame is due to a small number of very highly cited articles from 2007, 2010, and 2011. The 2007 paper *A short history of SHELX* (Acta Crystallogr. A 64, 112–122; 2008) by G.M. Sheldrick drew 9236 citations in the first 3 years after publication, and thus is a very special case. It is often used to discuss potential distortions due to skewed citation distributions and outliers. The other two papers received just over 500 citations each in their first 3 years and, while they drew considerable impact, they do not have the same distorting effect as the Sheldrick paper. Consequently, the RCI peak shown for Göttingen needs to be interpreted with

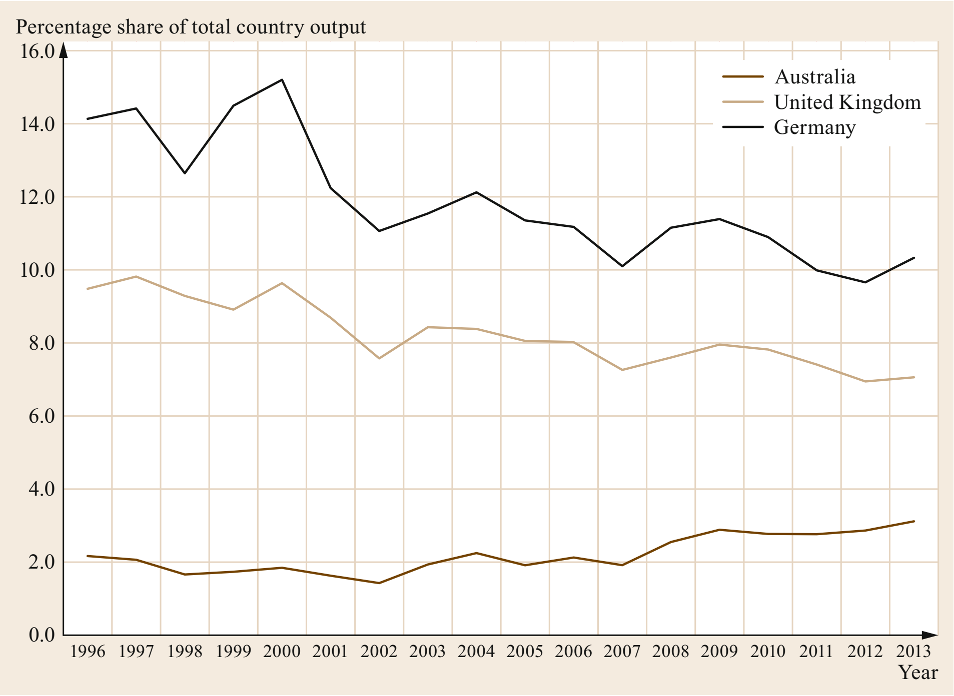


Fig. 18.6 Trends in the share of world publications in the top 10% most cited set for chemical sciences, selected countries, 1996–2013. Data source: Scopus, authors' calculations

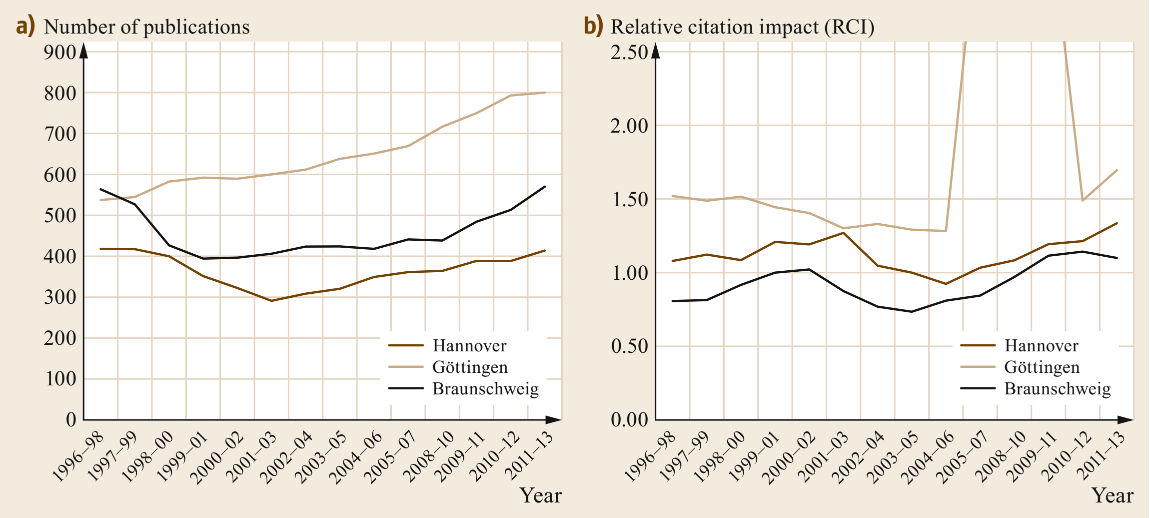


Fig. 18.7a,b Trends in publication output **(a)** and RCI **(b)** in the chemical sciences for selected German universities, 1996–2013. Data source: Scopus, authors' calculations

care. It does not reflect sustained performance change, but it does clearly demonstrate the rationale for using a suite of indicators rather than a single measure. In this case, complementary information based on size-independent or distribution-based indicators, as given in Table 18.2, are needed to correct for the distorting effects of outliers.

There is a very sharp increase in publication numbers for two of the three Australian universities we examined. While all three started from the same level of publication activity, Melbourne and, particularly, Monash show significant increases after 2000. However, the increase in numbers was not initially accompanied by similarly strong impact gains—a sustained

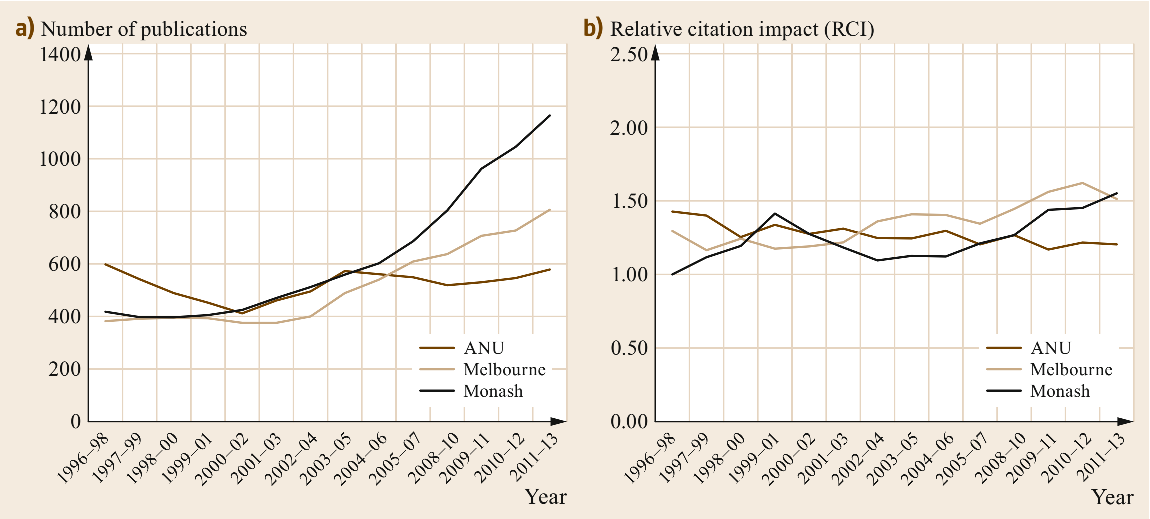


Fig. 18.8a,b Trends in publication output **(a)** and relative citation impact **(b)** in the chemical sciences for selected Australian universities, 1996–2013. Data source: Scopus, authors' calculations

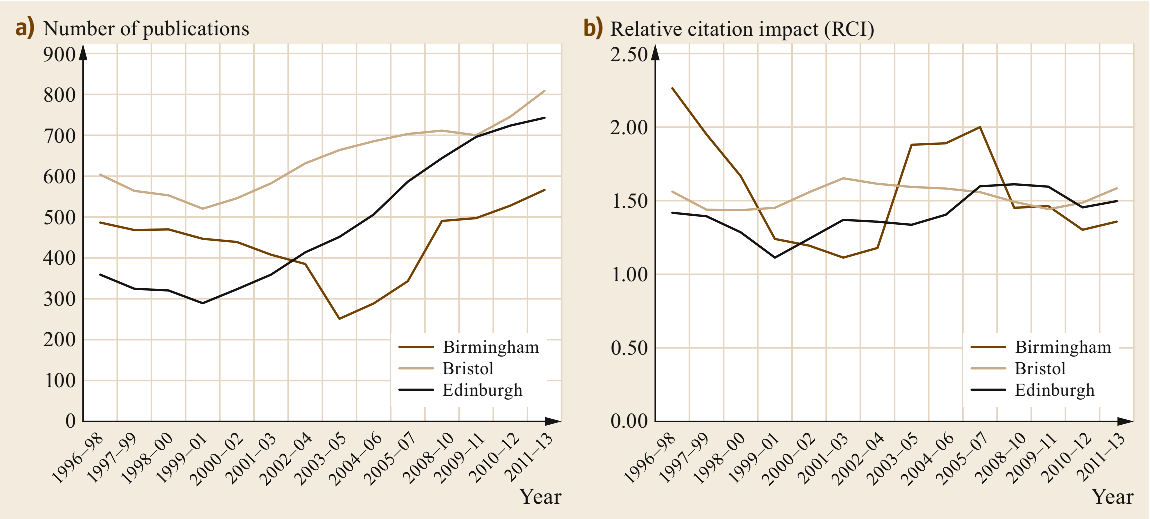


Fig. 18.9a,b Trends in publication output **(a)** and relative citation impact **(b)** in the chemical sciences for selected United Kingdom universities, 1996–2013. Data source: Scopus, authors' calculations

improvement in the average RCI did not occur for either institution until the second half of the period.

Whereas German universities maintained their relative standing in the RCI trends, this was not the case for the Australian universities (Fig. 18.8). The ANU had the strongest outcomes in terms of both quantity and impact at the beginning of the observation period but lost its leading position with regard to both performance dimensions by the end. Whatever external drivers of research performance existed, the internal drivers would appear to have more power, particularly in the case of the ANU.

Three different outcomes emerge for the three United Kingdom universities (Fig. 18.9). First, Edinburgh saw a huge increase in output in the period, and with a modest increase in the average RCI following an initial fall at the beginning. Second, Bristol also saw increasing publication numbers on a similar trajectory to Edinburgh, though not as steep, but their average RCI remained relatively stable across the period. Third, Birmingham's output has dropped below that of the other United Kingdom universities, and its average RCI dropped sharply at the start of the period; while it again overtook Edinburgh and Bristol, this was not a sustain-

Table 18.2 Assessment scores and bibliometric performance indicators for selected universities, 2008–2013

University	Assessment rating	Number of publications	Citations per publication	Relative citation impact	Percentage of output in top 1%	Percentage of output in top 10%
Hannover	Excellent	817	8.29	1.24	2.0	13.2
Göttingen	Excellent	1588	10.49	1.58	2.3	19.3
Braunschweig	Excellent	1072	7.56	1.14	1.0	12.2
ANU	3.7	1135	8.05	1.22	1.1	12.6
Melbourne	4.1	1649	9.58	1.46	1.8	15.9
Monash	4.7	2422	9.86	1.50	1.8	19.1
Birmingham	4.0	1151	9.38	1.41	2.1	16.2
Bristol	4.3	1538	10.46	1.56	1.8	19.0
Edinburgh	4.0	1434	9.83	1.48	1.7	18.9

Data source: Scopus, authors' calculations

able recovery, as it again lost ground at the end of the observation period.

We noted the variation in performance between universities that did not necessarily align with overall country trends, so we sought to investigate whether, within countries, they aligned with the outcomes of their relevant assessment exercises. For both the United Kingdom and Australia, we arrived at an overall score by either calculating a weighted average of the scores given for each discipline in the chemical sciences (Australia) or by weighting the distribution of scores given to outputs (United Kingdom). While the scores appear to be similar, they are not directly comparable (see Table 18.1 for a description of the different ratings). In Lower Saxony, the rating given was descriptive rather than numeric. The period analyzed in Table 18.2 closely aligns to the publication window used in all three assessments.

For Australia, the citation indicators appear to correlate well with the assessments given. This is not surprising given the central role that metrics played in the ERA process. In the United Kingdom, the results are not quite as clear cut. The university with the highest apparent rating (Bristol) outperforms the others on the two measures based on averages, but does not do so as convincingly on those based on the most highly cited publications in the field. For Germany, there is no differentiation in the assessment, though the data shows that Göttingen has a significantly stronger performance in all measures for this period (note: this period does not include the very highly cited paper by Sheldrick).

18.8.4 Institution Level—REF in Detail

We looked more closely at the output of United Kingdom universities (Table 18.3) as we were able to compare all output in the chemical sciences with that portion of the output that was submitted for assessment in REF2014 (i. e., the *best four* for each staff member).

Table 18.3 Bibliometric data for different sets of publications of the United Kingdom universities, 2008–2013

	Birmingham	Bristol	Edinburgh
Number of publications in chemistry journals:			
Submitted to panel 8	75	119	93
Submitted to all panels	205	186	184
Total for university	1145	1514	1427
Average relative citation impact			
Submitted to panel 8	2.19	2.75	2.76
Submitted to all panels	2.04	2.63	2.25
Total for university	1.41	1.57	1.48
Number of highly cited publications in chemistry journals (top 10%):			
Submitted to panel 8	24	57	47
Submitted to all panels	59	79	70
Total for university	183	289	270
% of highly cited publications in chemistry journals (top 10%):			
Submitted to panel 8	32	48	51
Submitted to all panels	29	42	38
Total for university	16	19	19

Data source: Scopus, authors' calculations

Table 18.4 Distribution of chemistry publications across REF 2014 panels

ERA Panel	Main Panel	Birmingham	Bristol	Edinburgh
Clinical Medicine	A	3	4	7
Allied Health Professions, Dentistry, Nursing, and Pharmacy	A	1	1	0
Psychology, Psychiatry, and Neuroscience	A	0	1	3
Biological Sciences	A	6	5	5
Agriculture, Veterinary and Food Science	A	0	1	1
Earth Systems and Environmental Sciences	B	10	0	5
Chemistry	B	36	67	49
Physics	B	4	14	4
Mathematical Sciences	B	0	2	1
Computer Science and Informatics	B	1	1	1
Aeronautical, Mechanical, Chemical, and Manufacturing Engineering	B	26	0	0
Electrical and Electronic Engineering, Metallurgy and Materials	B	11	0	0
Civil and Construction Engineering	B	0	0	0
General Engineering	B	0	6	25
Business and Management Studies	C	0	0	1
Total		100	100	100

Data source: Scopus, authors' calculations

Not unexpectedly, the performance of the submitted articles exceeds that for the total output of each university. That is only of concern if the rhetoric of universities and analysts then extrapolates from this relatively small sub-set of output to make declarations about the overall performance of all publications from their university in that discipline.

We undertook a supplementary analysis, looking at how many publications in our chemistry journals were submitted to panels other than the panel directly tasked with assessing the chemical sciences for the three case-study universities. These results are shown in Table 18.4.

Most publications were submitted within main panel B, but the distribution across panels within that varied significantly between universities. Two-thirds of outputs in the chemistry journal set that the University of Bristol submitted for assessment found their way to panel 8, but only just over one-third of outputs from Birmingham in the same set of journals were submitted to the same panel. At Birmingham, one-quarter were submitted to the engineering panel that included chemical engineering. Edinburgh submitted one-quarter of their outputs in the same journal to the general engineering panel.

We also looked at what proportion of the total output of the universities in the journal set were submitted for assessment and what proportion of highly cited articles from the same journal set were submitted. The results are shown in Table 18.5.

Table 18.5 Proportion of output in chemistry journals submitted to REF 2014

	Birmingham	Bristol	Edinburgh
Number of publications in chemistry journals:			
Submitted to panel 8	75	119	93
Submitted to all panels	205	186	184
Total for university	1145	1514	1427
% of total publications in chemistry journals:			
Submitted to panel 8	7	8	7
Submitted to all panels	18	12	13
Number of highly cited publications in chemistry journals (top 10%):			
Submitted to panel 8	24	57	47
Submitted to all panels	59	79	70
Total for university	183	289	270
% of highly cited publications in chemistry journals (top 10%):			
Submitted to panel 8	13	20	17
Submitted to all panels	32	27	26

Data source: Scopus, authors' calculations

There are differences in the definition of a university output between our data set and the REF data set. To be classified to a university in our data set, the institution must be listed in the addresses given for the publication. In REF, academics may *bring* publications with them if they have recently moved to the university, though the institution's name will not appear in the addresses for that article. Similarly, if an academic has left the university for a new appointment in the REF census period, the publication will be submitted to REF by the new institution. However, we feel that these two factors will cancel out and the analysis should remain robust.

In total, less than 10% of the journal publications we identified for each university were submitted for assessment to panel 8. In contrast, we find that a far higher proportion of highly cited publications were submitted. Were universities anticipating that citations might play a greater role in panel deliberations and thus made a deliberate decision to select more highly cited items? Or is it the case that the *best outputs* tended to attract much higher citations? Whatever the reason, it is clear that the submitted output is not necessarily representative of the whole output of an institution.

18.9 Discussion and Conclusions

While the three evaluation systems—in Australia, Germany, and the United Kingdom—examined here differ considerably, we do not find the anticipated differences in changing performances, either at the country level or at the institutional level. For all three countries, we observe continuously increasing publication numbers since 2002. At the national level, the different trends in relative citation impact for the chemical sciences are more pronounced. All three countries, however, still remain well above the world average on this measure. Australia's improved performance was more pronounced than was the case for the United Kingdom, and Germany's performance even decreased slightly. This is also reflected in the increasing share of highly cited publications (top 10%) for Australia, while the German and United Kingdom shares remain stable. However, for all countries, the share of highly cited publications is still well above the world average.

When it comes to the institutional level, the picture is less clear. In each of the three countries, and also for most of our case study institutions, we see improved performance. This does not permit us to draw any general conclusions about the effectiveness of any of the systems in place. If we were to base our assessment on these systems using indicators alone, the data would not allow us to conclude that any of the systems has a beneficial or detrimental influence on performance.

Thus, it appears that it is not the specific system that makes a difference but, rather, the fact that performance becomes a central topic of conversation. Consequently, universities pay increasing attention to monitoring performance—be it based on quantitative, qualitative, or mixed information. Global developments such as international university rankings also play a role in heightening awareness of performance.

In order to better understand the mechanisms behind changing performance, restricting scrutiny to mere numbers is insufficient. Contextual information at various levels of aggregation—within and outside the institutions—is highly relevant. This is a requirement repeatedly formulated by bibliometricians when it comes to applying research metrics.

References

18.1 J. Wilsdon, L. Allen, E. Belfiore, P. Campbell, S. Curry, S. Hill, R. Jones, R. Kain, S. Kerridge, M. Thelwall, J. Tinkler, I. Viney, P. Wouters, J. Hill, B. Johnson: *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management* (Higher Education Funding Council for England, Bristol 2015)

18.2 P. Jump: Evolution of the REF, https://www.timeshighereducation.com/features/evolution-of-the-ref/2008100.article (2013)

18.3 HEFCE: REF 2014 Key Facts, http://www.ref.ac.uk/media/ref/content/pub/REF%20Brief%20Guide%202014.pdf (2014)

18.4 P. Cook: Science and Technology Budget Statement 1995/1996, https://www.industry.gov.au/sites/g/files/net3906/f/1995-96-science-technology-budget-statement.pdf (2014)

18.5 R.D. Linke: *Performance Indicators in Higher Education* (Australian Govt. Pub. Service, Canberra 1991)

18.6 D.A. Kemp: Knowledge and innovation: A policy statement on research and research training, http://www.voced.edu.au/content/ngv:41605 (1995)

18.7 Universities Australia: *Higher Education Research Facts and Figures* (Universities Australia, Canberra 2015)

18.8 The Australian Research Council: ERA Discipline Matrix, https://www.arc.gov.au/excellence-research-australia/archived-era-documents
18.9 The Australian Research Council: Excellence in Research for Australia, ERA 2015 Submission Guidelines (Commonwealth, Australia 2014)
18.10 Wissenschaftsrat: Recommendations for rankings in the system of higher education and research. Part 1: Research, https://www.wissenschaftsrat.de/download/archiv/6285-04-engl.pdf (2004)
18.11 Wissenschaftsrat: Empfehlungen zur Zukunft des Forschungsratings, https://www.wissenschaftsrat.de/download/archiv/3409-13.pdf (2013)
18.12 Wissenschaftsrat: Empfehlungen zu einem Kerndatensatz Forschung, https://www.wissenschaftsrat.de/download/archiv/2855-13.pdf (2013)
18.13 Wissenschaftsrat: *Empfehlungen zur Spezifikation des Kerndatensatz Forschung* (Wissenschaftsrat, Berlin 2016)
18.14 S. Biesenbender, S. Hornbostel: The research core dataset for the German science system: Developing standards for an integrated management of research information, Scientometrics **108**(1), 401–412 (2016)
18.15 Wissenschaftliche Kommission Niedersachsen: *Forschungsevaluation an niedersächsischen Hochschulen und Forschungseinrichtungen. Grundzüge des Verfahrens* (Wissenschaftliche Kommission Niedersachsen, Hannover 2004)
18.16 Wissenschaftliche Kommission Niedersachsen: http://www.wk.niedersachsen.de/kommission/aufgaben/aufgaben-72387.html
18.17 Wissenschaftliche Kommission Niedersachsen: http://www.wk.niedersachsen.de/taetigkeitsbereiche/forschungs_und_strukturevaluation/faecherbezogene_forschungsevaluation/evaluationsberichte-137324.html
18.18 Wissenschaftliche Kommission Niedersachsen: http://www.wk.niedersachsen.de/taetigkeitsbereiche/forschungs_und_strukturevaluation/themengeleitete_evaluation/evaluationsberichte-137331.html
18.19 U. Schmoch, S. Gruber, R. Frietsch: 5. Indikatorbericht Bibliometrische Indikatoren für den PFI Monitoring Bericht 2016, Institut für Forschungsinformation und Qualitätssicherung (iFQ), Fraunhofer-Institut für System- und Innovationsforschung ISI, Universität Bielefeld, Institute for Interdisciplinary Studies of Science (I^2SoS) https://www.bmbf.de/files/5.%20Indikatorbericht_PFI_Monitoring_Bericht_2016-03-25.pdf (2016)
18.20 HEFCE: REF 2014 Assessment Framework and Guidance on Submissions: REF 02.2011, https://www.ref.ac.uk/2014/media/ref/content/pub/assessmentframeworkandguidanceonsubmissions/GOS%20including%20addendum.pdf (2011)
18.21 Australian Bureau of Statistics: 1297.0 – Australian and New Zealand Standard Research Classification (ANZSRC), http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/2A3A6DB3F4180D03CA25741A000E25F3/$File/12970_2008.pdf (2008)
18.22 HEFCE: REF 2014, Panel Criteria and Working Methods, Main Panel B Criteria, https://www.ref.ac.uk/2014/media/ref/content/pub/panelcriteriaandworkingmethods/01_12_2B.pdf (2012)
18.23 The Australian Research Council: ERA 2015 Evaluation Handbook, https://www.arc.gov.au/excellence-research-australia/archived-era-documents (2017)
18.24 Wissenschaftliche Kommission Niedersachsen: *Forschungsevaluation Chemie 2015: Ergebnisbericht* (Wissenschaftliche Kommission Niedersachsen, Hannover 2015)
18.25 HEFCE: 2014 REF Results and submissions, https://results.ref.ac.uk/(S(rnp10trrqqx3l5icc2wr3oeo))/ (2014)

Sybille Hinze

Research System and Science Dynamics
German Center for Higher Education Research and Science Studies (DZHW)
Berlin, Germany
hinze@dzhw.eu

Sybille Hinze is Deputy Head of the Research System and Science Dynamics Department at the German Center for Higher Education Research and Science Studies. She graduated in Management of Science from Humboldt-University in 1990 and received her PhD from Leiden University in 1997. She worked at Fraunhofer Institute for Systems and Innovation Research and Australian National University. Her research focuses on research performance and structures of research and innovation systems.

Linda Butler

School of Politics and International Relations
The Australian National University
Acton, Australia
linda.butler52@gmail.com

Linda Butler is a Visitor at the School of Politics and International Relations at The Australian National University in Canberra. Prior to retiring, she was Head of the Research Evaluation and Policy Project at the ANU. She has been a researcher in bibliometrics for over 20 years, more recently with a focus on their application in national research assessment exercises, and their extension to the Arts and Humanities.

Part C | 18

Paul Donner

Research Area 2, Research System and Science Dynamics
German Centre for Higher Education Research and Science Studies, DZHW
Berlin, Germany
donner@dzhw.eu

Paul Donner is a scientometrics researcher at the German Centre for Higher Education Research and Science Studies, Research Area 2, Research System and Science Dynamics. He obtained a Master of Arts degree in Library and Information Science at Humboldt University Berlin. His work includes bibliometric research, primarily into indicator validity, bibliometric contributions to contracted science evaluation studies and the quality assurance of bibliometric research data infrastructure.

Ian McAllister

School of Politics and International Relations
The Australian National University
Acton, Australia
ian.mcallister@anu.edu.au

Ian McAllister is Distinguished Professor of Political Science at The Australian National University. He previously held chairs at ADFA and the University of Manchester. He has published several books. He has been Director of the Australian Election Study since 1987, a large national post-election survey of political attitudes and behaviour. His research covers Australian politics, comparative political behaviour and post-Communist politics. He is currently completing a book on Russian voting.

19. Scientific Collaboration Among BRICS: Trends and Priority Areas

Jacqueline Leta, Raymundo das Neves Machado, Roberto Mario Lovón Canchumani

The political and economic partnership known as BRIC (for Brazil, Russia, India and China) was formally established in 2008. Three years later, in a joint meeting in Cape Town, a new member, South Africa, was included in the group. In this meeting, BRICS delegates elaborated a list of priority areas for enhancing bi- or multilateral cooperation in the fields of science, technology and innovation. Considering the growing importance of BRICS in the global economy and other sectors, the present study investigates the performance of the group in the scientific arena before and after its formalization in 2008, looking closely at BRICS collaborative publications, in order to identify whether the priority areas established in the Cape Town declaration are being actually pursued. Data were collected during February and March 2017 from the Web of Science database, covering the period 2000–2015. To match scientific collaborations, specific searches were carried out by combining the names of two BRICS members and time periods. Various bibliometric techniques were used, including diachronic analysis, Bradford's law and journal co-citation analysis. Among the key findings highlighted here are a marked increase in BRICS participation during the period, widely varying levels of collaboration among members, and the presence of physics as a central field for most members. The chapter concludes with an in-depth discussion focusing on correlations between the fields with greater collaboration and the priority areas.

19.1 BRICS: From Origin to Priority Areas in ST&I

The beginning of the twenty-first century was marked by a severe economic crisis that affected the major world powers. This scenario has favored the formation of new geopolitical configurations, such as the BRIC, the acronym for the partnership between Brazil, Russia, India and China. The BRIC concept was introduced in 2001 by Jim O'Neill, who predicted that GDP growth among the BRIC would soon exceed that of the G7 countries [19.1].

At that time, the BRIC comprised the largest peripheral countries, including not only the two largest in population but also the two oldest. In addition, the group constituted almost half of the world's workforce, held a huge bulk of the global natural resources and sported booming economies. Other features that favored the establishment of the group as a new economic and political alliance included the expanded purchasing power of peripheral countries, especially China and South American countries such as Brazil, and the aggressive stance of Brazil and India towards the liberalization of world trade, especially of agricultural goods [19.2].

The BRIC emerged as an alternative to the traditional north–south model of cooperation, which at the time was "experiencing depletion" [19.3, p. 7]. Together, Brazil, Russia, India and China could have greater representation in global governance, and thus could broaden their participation in the power arena,

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_19

restricted at that time to the economic and political core countries known as the G7. *Cassiolato* and *Lastres* [19.4] highlighted various socioeconomic conditions that favored the creation of the BRIC alliance, such as their strategic geoposition, their diversity and rich territory with huge populations, and their economic performance, including growth in their export and import levels as well as in their foreign direct investments. The authors also highlighted that the 2007–2008 global economic crisis had "repositioned the role and importance of the BRIC" [19.4, p. 16]; in other words, the economic crisis propelled BRIC to occupy a more central position in the global economy.

The BRIC group was formally established during a meeting of the countries' foreign affairs ministers hosted in Yekaterinburg, Russia, in 2008. Since then, the group has organized annual meetings among the presidents or prime ministers of all the BRIC countries. The aim of these summits was both to strengthen cultural and political ties and to develop a common agenda of initiatives and actions to empower and consolidate the group. During the 2011 Sanya meeting in China, a new partner—South Africa—was included in the group, and an *S* was added to the acronym, officially establishing the BRICS group. In this meeting, the five countries discussed strategies to enhance their economies and to strengthen their ties in various sectors, including agriculture, energy, and science and technology, launching the first action plan [19.5].

In 2014, for the first time, BRICS delegates in science, technology and innovation (ST&I) met in Cape Town, South Africa, to discuss strategies to enhance their cooperation in this sector. The group established a list of priority areas in which the countries should initiate or increase bi- or multilateral cooperation. The list included [19.6]:

- Innovation and technology transfer
- Food security and sustainable agriculture
- Climate change and natural disaster preparedness and mitigation
- New and renewable energy, energy efficiency
- Nanotechnology
- High-performance computing
- Basic research
- Space research and exploration, aeronautics, astronomy and Earth observation
- Medicine and biotechnology
- Biomedicine and life sciences (biomedical engineering, bioinformatics, biomaterials)
- Water resources and pollution treatment
- High-tech zones/science parks and incubators
- Technology transfer
- Science popularization
- Information and communication technology
- Clean coal technologies
- Natural gas and nonconventional gases
- Ocean and polar sciences
- Geospatial technologies and their applications.

Many of these priority areas embrace a number of hot and emerging hot topics [19.7, 8], as they are also in accordance with the OECD 2016 report, which analyzes disruptive and promising trends in science, technology and innovation not only for OECD countries, but also for major nonmember OECD countries, including all BRICS countries [19.9].

Considering BRICS as a political-economic union of five large countries that have been capturing space in global governance, the present study aims to investigate how this group has performed in the scientific arena, before and after its formalization in 2008, looking closely at thematic trends among BRICS collaborative publications.

19.1.1 BRICS and Bibliometric Studies

The BRICS contribution to global science has been fairly well investigated, including studies with different approaches and/or the use of multiple indicators in science, technology and/or innovation [19.10–15].

In the following, we present a brief overview of bibliometric studies investigating trends in BRIC or BRICS scientific publications.

The bibliometric literature focusing on BRICS' scientific publications includes those with a more general approach, such as the study by *Wagner* and *Wong* [19.16], who investigated BRIC representation in the Science Citation Index Expanded (SCIE). Based on national and international sources, the authors compiled a list of 15 000 titles of BRIC national journals, of which 445 titles (almost 3%) were indexed in SCIE. According to the authors, this share is similar to that found in other sources. Considering such similarities, Wagner and Wong argued that the BRIC countries were not underrepresented in SCIE, but did note that many BRIC publications were less visible than those published in European or North American titles.

BRICS scientific excellence has also been examined based on their presence in the most frequently cited papers, that is, the top 1% and top 10% [19.17]. The authors found that China exhibited outstanding performance, but that, as with other BRICS countries, the share of Chinese publications (0.7 and 7.7%) among the 1 and 10% top-cited papers was below expectations. South Africa, Brazil, India and Russia, in descending order, were less represented among the set of top-cited papers. Focusing on citable documents reported by SCImago, *Guevara* and *Mendoza* [19.18] investigated the similarities and differences in publishing patterns among BRIC members. The authors used accurate statistics to measure the relative importance of thematic categories (based on journal classification) for each country. They found that BRIC members as a whole were more dissimilar, tending to be closer in terms of publishing patterns to their neighbors or to countries that shared the same language; India, however, was the exception.

The scientific collaboration involving BRICS countries is also a frequent theme in bibliometric literature. *Finardi* [19.19], for example, analyzed collaboration based on co-authorships of BRICS publications published during the period from 1980 to 2012. The results indicated that collaboration among the five BRICS has grown, especially after the 2000s. For the whole period, the strongest linkages were found between India and South Africa and between Brazil and South Africa. Engineering was the area with the highest share of collaboration in the former pair of countries, whereas medical sciences was highest in the latter. *Singh* and *Hasan* [19.20] investigated BRICS research output, including partners in publications, over three decades (1994–2013). According to the authors, around 26% of BRICS publications represented collaborations with another country, most frequently the USA, followed by Germany, England, Japan and France. Using probabilistic affinity indexes, *Finardi* and *Buratti* [19.21] also investigated key patterns in BRICS scientific publications, analyzing the group's linkage with 65 other countries. Geographical, cultural and historical proximity were found to be the major factors driving collaboration between BRICS and other countries.

Other studies have explored BRICS' contribution in specific scientific fields or areas. *Bai* et al. [19.22] analyzed the contribution of the group in literature on neglected diseases, using PubMed as the main informational source. The results revealed that Brazil, China and India were among the ten countries with the largest share of articles in this field. Using cluster analysis, the authors found eight groups of words, each associated with a neglected disease. Helminthiasis, human immunodeficiency virus/acquired immune deficiency syndrome (HIV/AIDS) and tuberculosis were the most frequent diseases among the BRICS. Other examples include the work on stem cells, led by Machado. The first studies focused on the performance of Brazilian scientists in this field [19.23, 24]. In later work, the authors turned their attention to mapping author productivity and main thematic categories for BRICS publications on stem cells [19.25]. Their results showed an increase in the number of authors with higher levels of productivity in this field across all BRICS countries. In recent years, BRICS' research on stem cells has reflected a strong emphasis not only in clinical medicine, especially hematology and oncology, but also in areas such as cell and molecular biology. According to the authors, this picture suggests a shift in the orientation of BRICS' publications on stem cells towards an understanding of stem cell mechanisms of division and differentiation.

Yang et al. [19.26] explored the disciplinary structure of G7 and BRIC publications during three separate years (1991, 2000 and 2009). The authors found a marked change in disciplinary structure among the BRIC, with a focus on basic research in physics, chemistry, mathematics and engineering. In contrast, the G7 publications were oriented more towards life sciences. Despite such differences, authors noted that the BRIC disciplinary structure has become more similar to that of G7 over time.

The above-mentioned studies represent a small fraction of the huge body of literature on BRICS' performance in science, including trends in journals, citation, collaboration, and specific fields and disciplines. The present study offers new insight into BRICS' performance in science, as it investigates the intellectual or disciplinary structure of BRICS' collaborative publications in order to identify whether the main themes addressed by BRICS in this set of publications is in accordance with the priority areas established in the Cape Town declaration.

19.2 Methodology

General data for the BRICS' scientific publications were collected directly from the *Web of Science* (WoS) core collection (Clarivate Analytics) database in March 2017 and analyzed using Microsoft Excel software. Descriptive analyses based on this data set are presented in an introductory section, including analysis of the compound annual growth rate (CAGR), which was calculated as follows

$$\mathrm{CAGR}\,(t_0, t_n) = \left(\frac{V(t_n)}{V(t_0)}\right)^{\frac{1}{t_n - t_0}} - 1\,,$$

where $V(t_0)$ = start value, $V(t_n)$ = finish value and $t_n - t_0$ = number of years.

BRICS collaborative publications were downloaded from the *WoS* (Clarivate Analytics) during February and March 2017 by combining the names of two of the BRICS countries (for instance, Brazil and Russia) in the *address* filter and the period 2000–2007 or 2008–2015 in the *year of publication* filter. Downloading was limited to 500 publications at a time, and each archive included complete information for all publications, including references, in a CSV (comma-separated values) format. These archives were then imported into Microsoft Excel software in order to identify inconsistencies and to exclude duplicates. Based on this set of publications, two main analyses were carried out, Bradford analysis and journal co-citation analysis, for each country in both periods.

Bradford's statistical model [19.27] was applied to identify core journals, that is, those in zone 1. Although fewer titles are included in Bradford's first zone, it comprises, in general, the largest number of articles, in this case, co-authored by two or more BRICS countries. Bradford statistics on BRICS' collaborative publications were calculated using Microsoft Excel.

The journal co-citation analysis as described by *McCain* [19.28] reveals the intellectual structure, or the main domains, of a set of publications. Considering the premise that the greater the frequency of co-cited pairs, the closer the relationship between the pairs, journal co-citation analysis allows one to identify clusters of journals with thematic similarities. Co-citation analysis on BRICS collaborative publications was carried out with the help of Gephi, an open-source software program for network analysis developed in 2009 [19.29]. Ten maps based on journal co-citation, one for each of the five BRICS countries during the periods 2000–2007 and 2008–20015, were then generated with Gephi. The analysis considered only journals that had been cited more than 100 times during a period. The identification of journal communities (or clusters) used the algorithm elaborated by *Blondel* et al. [19.30], in which nodes (journals) are grouped in neighborhoods with the greatest contribution to the positive variation in the modularity measure. Fruchterman-Reingold's classical spatial distribution algorithm, available in Gephi, was used to gain better visualization of the clusters; labels were also adjusted for better visualization of the nodes.

19.3 Results

The results are presented in three main sections: general trends in BRICS scientific output, trends in BRICS collaborative scientific publications, and scientific structure in BRICS collaborative scientific publications.

19.3.1 BRICS Scientific Production: General Trends

From 2000 to 2015, BRICS published 6 520 943 scientific publications. China was responsible for 61.03% of all publications, India for 16.47%, Russia for 10.08%, Brazil for 9.74% and South Africa for 0.68%. Among the total publications from BRICS during the period, the most frequent types were articles (74.06%) and conference papers (20.47%). Reviews and other types of publications summed to less than 6%.

The annual number of scientific publications for each of the BRICS member countries is shown in Fig. 19.1. A strong exponential growth trend ($r^2 > 0.90$, p-value < 0.05) can be observed. During the period, China exhibited the smallest rate of duplication time (2.5 years), while Russia had the largest (4.1 years); in other words, China had the fastest growth and Russia the slowest.

The inset in Fig. 19.1 shows the increase in the number of publications for each country over two periods, 2000–2007 and 2008–2015, representing the 8 year periods before and after the formal establishment of BRICS.

A better view of this increase can be seen in Table 19.1, which presents the total number of publications for each period and the compound annual growth rate (CAGR). The number of BRICS scientific publications demonstrates impressive growth from 2000–2007 to 2008–2015, with a CAGR of 15.70. Among the BRICS countries, China exhibits the highest CAGR, whereas Russia shows the lowest. This result is in accordance with the rates of duplication times indicated previously for China and Russia.

Table 19.1 The growth in BRICS publications indexed in the WoS database

Country	2000–2007	2008–2015	CAGR[a] 2000–2015
Brazil	186 061	449 372	13.42
Russia	283 151	374 307	4.07
India	284 302	789 418	15.71
China	919 956	3 059 658	18.73
South Africa	53 919	120 799	12.21
BRICS	1 727 389	4 793 554	15.70

[a] CAGR = compound annual growth rate

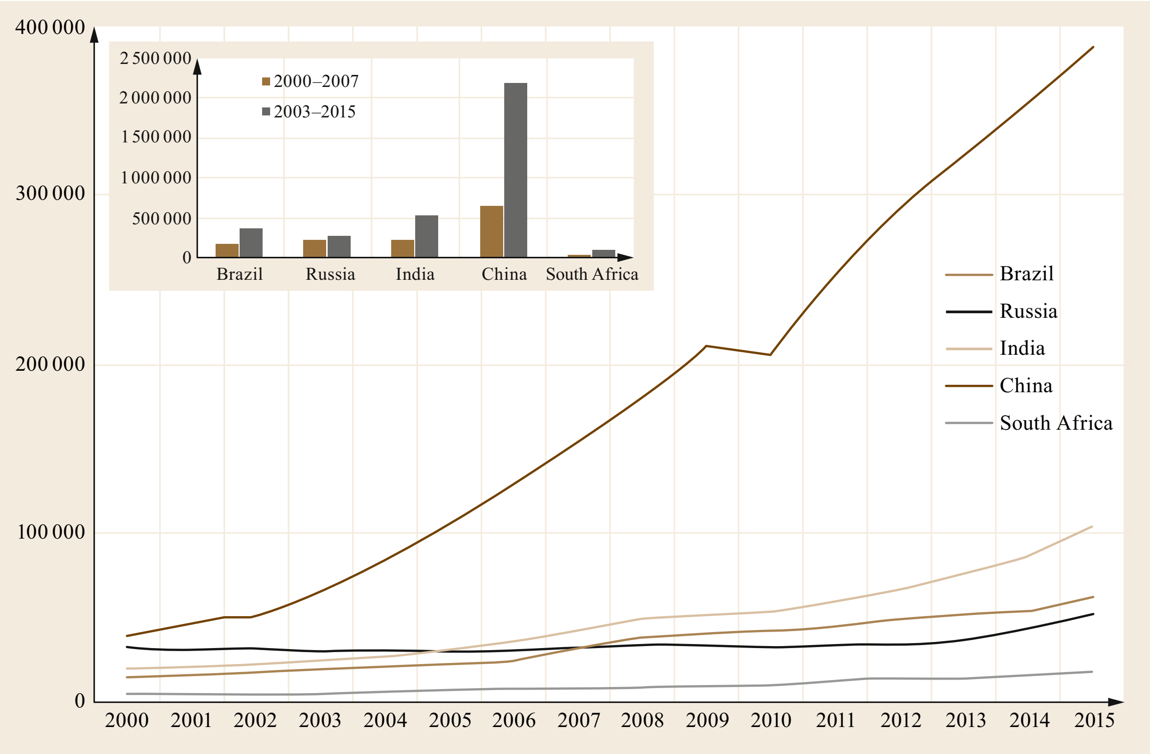


Fig. 19.1 BRICS' annual number of scientific publications in WoS database from 2000 to 2015

Wong and *Wang* [19.14] studied the growth trajectory of BRICS publications and observed three growth phases: slow growth in the initial years, fast exponential growth in the middle years, and no growth (or steady state) in later years. The authors note that China is experiencing the highest rate of growth, in the middle of the second growth phase. Russia, on the other hand, is already in the third phase, and thus entering a period of maturity, with little or no growth. Based on a logistic growth function analysis, the authors predicted that China will enter the mature phase when the country reaches around 3.3 million publications, whereas the other BRIC members will reach maturity at numbers below one million.

The growth observed in the other BRICS countries may be influenced by different factors, including the well-known expansion of the WoS/Clarivate Analytics database. *Testa* [19.31] recently presented details regarding the increased number of journals indexed in WoS across ten countries, including Brazil, India, China and South Africa. The author posited that the motivation behind such an initiative was the desire to enhance visibility of regional journals and "their local importance" [19.31, p. 100]. Regarding database journal coverage, *Leta* [19.32] analyzed Brazilian growth in terms of the number of publications indexed in two international databases, including WoS. In the author's view, the increase was a result not only of internal factors such as an increase in science (S) and technology (T) resources and qualified human resources, but also of external factors, including the growth in the total number of Brazilian titles indexed in the databases.

There is no doubt that both internal and external factors have contributed to China's remarkable growth in terms of the number of scientific publications. Using data retrieved from the Ulrichsweb database, *Wang* et al. [19.33] presented a series of results revealing that China, which is among the top three publishing countries, experienced fluctuations in scholarly publishing over the period 1950–2013. According to the authors, such variation was related to the establishment of new internal rules, the Interim Regulations on the Administration of Periodical Publications, which to a certain extent prioritized quality over quantity of journals. Additional factors may also contribute, such as the search for new sources of publishing. *Yan* et al. [19.34], for instance, examined China's contributions in *PLOS One*, one of the largest and most successful open-access journals in the world, and observed exponential growth in the number of Chinese publications in this journal. In fact, in 2013, more than 6000 papers published in *PLOS One* had a Chinese author, accounting for almost 20% of its total volume that year.

Such a strong presence for China in a single open-access journal raises the possibility that this new paradigm in scientific communication also drove the growth in publications among the other BRICS countries. Hence, it seemed relevant to investigate how large a proportion of the BRICS articles were published in open-access journals. Table 19.2 shows the number of articles published in open- and restricted-access journals for each of the five BRICS countries. From 2000–2007 to 2008–2015, the five countries increased the number and share of articles in open-access journals. In both periods, China had the largest number of such publications, while South Africa had the lowest. Brazil displayed the highest share of articles in open-access journals in both periods (10.36 and 27.04%), while Russia exhibited the lowest (0.19 and 5.58%). Despite having the lowest share, however, Russia demonstrated the most impressive growth in this type of publication from one period to the other.

Comparing Tables 19.1 and 19.2, one can surmise that scientists in BRICS have invested in publishing in open-access journals as a strategy to gain visibility in the global science community. Despite the lack of familiarity and the criticism from scientists regarding this publishing model [19.35], there is considerable evidence to suggest that this type of publication does have quality and may contribute to increasing a paper's visibility and citation frequency [19.36, 37]. Such impact may be the reason behind BRICS' huge investment in developing research-oriented repositories as a strategy for disseminating scientific knowledge, as described by *Calderón-Martïnez* and *Ruiz-Conde* [19.11].

As a final analysis, Table 19.3 shows the fields in which BRICS publications most frequently appear. In 2000–2007, with the exception of South Africa, BRICS publications were strongly oriented towards the exact sciences and engineering fields. When Chinese publications are considered, nine of the most prolific fields fall into these categories. Brazilian publications were half dispersed in the exact sciences, while those for South African were spread among categories. The main fields for South African publications differed from the rest of BRICS, including fields in the biological sciences, such as environmental and plant science and zoology, which are closely aligned with the country's local issues, including its diverse ecosystems, flora and fauna. Brazil, which has one of the world's largest forests, exhibited a more international profile, while South African publications emphasized local issues that drove the country's mainstream research.

For the period 2008–2015, publications from China, Russia, India and Brazil, in descending order, again displayed a strong orientation towards exact sciences and engineering. This trend was still strongest in China.

For Brazilian publications, two new fields, pharmacology and environmental, science ecology, emerged in 2008–2015, replacing dentistry and neurosciences. This movement may signal a new orientation of Brazilian research towards local issues related to the country's fauna and flora, especially in Amazonia, which has the greatest biodiversity in the world.

Considering South African publications, it is clear that the country's mainstream research prioritizes local issues. In this period, a movement is also observed: infectious disease and environmental, science ecology replacing zoology and agriculture. Such growing interest in research on infectious disease in more recent years may be a response to a persistent scenario on the African continent, where hundreds of millions of people are victims of diseases such as malaria, yellow fever, dengue and HIV/AIDS. In recent years, various initiatives have been implemented, including the establishment of the South African Global Disease Detection (GDD) Center in 2010, to improve research in biomedicine and health sciences, including emerging infectious diseases [19.38].

Table 19.2 BRICS' scientific articles in open- or restricted-access journals indexed in the WoS database

	Brazil	Russia	India	China	South Africa
2000–2007					
Open access	22 492	4720	12 697	8168	2488
Restricted access	145 878	241 915	210 481	664 644	46 755
Total	168 370	246 635	223 178	672 812	49 243
% Open access	13.36	0.19	5.69	1.21	5.05
2008–2015					
Open access	104 218	16 863	54 659	151 649	14 673
Restricted access	281 158	285 351	491 151	2 057 864	90 539
Total	385 376	302 214	545 810	2 209 513	105 212
% Open access	27.04	5.58	10.01	6.86	13.95

Table 19.3 The ten most prolific fields of BRICS' publications, indexed in the WoS database

	Brazil	Russia	India	China	South Africa
2000–2007					
1	Physics	Physics	Chemistry	Engineering	Engineering
2	Chemistry	Chemistry	Physics	Chemistry	Environmental science ecology
3	Engineering	Engineering	Engineering	Computer science	Chemistry
4	Agriculture	Materials science	Materials science	Physics	General internal med
5	Computer science	Optics	Computer science	Materials science	Plant science
6	Biochem mol biology	Mathematics	Agriculture	Mathematics	Physics
7	Dentistry oral surg med	Astron and astrophysics	S and T other topics	Optics	Computer science
8	Neurosc neurology	Biochem mol biology	Biochem mol biology	Telecommunications	S and T other topics
9	Materials science	Instruments instrumentation	Pharmacology pharmacy	Autom control systems	Zoology
10	Veterinary science	Geology	Mathematics	Biochem mol biology	Agriculture
2008–2015					
1	Engineering	Physics	Engineering	Engineering	Engineering
2	Agriculture	Chemistry	Chemistry	Materials science	Environmental science ecology
3	Physics	Engineering	Physics	Chemistry	Chemistry
4	Chemistry	Materials science	Computer science	Computer science	Physics
5	Computer science	Mathematics	Materials science	Physics	Plant science
6	Biochem mol biology	Optics	S and T other topics	S and T other topics	Infectious diseases
7	Neurosc neurology	Astron and astrophysics	Pharmacology pharmacy	Mathematics	S and T other topics
8	Materials science	Biochem mol biology	Biochem mol biology	Automat control systems	General internal med
9	Pharmacology pharmacy	Geology	Telecommunications	Biochem mol biology	Computer science
10	Pub environment and science ecology	S and T other topics	Agriculture	Optics	Pub environment and science ecology

19.3.2 BRICS Collaborative Articles: Main Journals

As shown previously, with the exception of South Africa, BRICS scientific publications are mainly oriented towards the fields of exact sciences and engineering. In order to delve more deeply into this information, we turn our attention to an analysis of the fields and main thematic content of BRICS' collaborative articles, that is, publications co-authored by two or more BRICS countries.

As an initial step, Bradford analysis was applied to the journals in which the collaborative articles were published. Table 19.4 shows the total number of journals and articles as well as the number of zones found for the collaborative articles of each country in both pe-

Table 19.4 Journal distribution of BRICS collaborative articles indexed in the WoS, according to Bradford's model

Country	Total		Number of zones	$r^{2\ a}$	Zone 1	
	Journals	Articles			Journals	Articles
2000–2017						
Brazil	943	2944	4	0.96	9	845
Russia	1155	5916	3	0.85	12	2338
India	832	2601	3	0.96	9	951
China	1400	4681	4	0.93	10	1410
South Africa	598	1053	3	0.85	5	79
2008–2015						
Brazil	1228	3470	4	0.93	8	758
Russia	1271	5196	4	0.97	10	1676
India	1329	3973	4	0.93	8	975
China	1993	6142	5	0.90	11	1206
South Africa	994	2137	6	0.87	6	181

[a] Coefficient of determination

riods, along with the coefficient of determination (r^2), which indicates that all data had a good fit to the Bradford distribution model ($r^2 > 0.80$).

As a first observation, the increase in BRICS' collaborative articles varied considerably, from negative growth observed for Russian articles to 103% growth for South Africa. China showed the largest number of co-authored papers, displaying a 31.2% increase during the period.

Bradford analysis for BRICS' collaborative articles in 2000–2007 showed that the number of journals varied from 598 (South Africa) to 1400 (China). This set of journals was distributed across three or four zones. As for the core journals, that is, journals found in zone 1, Russia had the largest nucleus, comprising 12 titles that jointly published 2338 articles, almost 40% of the country's total articles during this period. South Africa had the smallest number of core journals, with five titles, in which 79 (less than 8%) collaborative articles were published.

For the period 2008–2015, Bradford analysis revealed a different model for journal distribution, with the number of zones varying from four to six; Brazil was the only country in which the number of zones remained the same as in the previous period. With regard to the core journals in zone 1, South Africa again displayed the smallest nucleus, while China exhibited the largest, with 11 titles encompassing 1206 articles, almost 20% of all Chinese collaborative publications with a BRICS partner.

Table 19.5 lists journals included in zone 1 for each BRICS country in the period 2000–2007. Upon first observation, a high overlapping of journals can be seen, mainly related to the exact sciences, specifically in physics. Among the top-ranked core journals, six appeared in the Brazilian, Russian, Indian and Chinese lists, including *Physical Review Letters*, *Physics Letters B*, *Physical Review C*, *Physical Review D* and *European Physical Journal C*. It is important to highlight that with the exception of *Physical Review Letters*—a journal with a broader spectrum of subjects—the set of journals shared by BRICS included specific themes such as condensed matter and materials physics (*Physics Letters B*), nuclear physics (*Physical Review C*) and particle physics, gravitation and cosmology (*Physical Review D*).

A different trend, however, is observed for South Africa, encompassing a broader list of journals in different fields, including, for instance, *Monthly Notices of the Royal Astronomical Society* in physics, the *International Journal of Systematic and Evolutionary Microbiology* in biomedicine, and *The Lancet* in medicine.

Table 19.6 presents a list of journals from zone 1 for each BRICS country during the period 2008–2015. One general observation is that the number of journals was reduced compared to the previous period.

Table 19.5 List of core journals of BRICS collaborative articles, indexed in the WoS, 2000–2007

Full name	Abbreviation	Brazil	Russia	India	China	South Africa	Total
Physical Review Letters	*Phys Rev Lett*	236	703	340	510	–	1789
Physical Review D	*Phys Rev D*	128	474	198	325	–	1125
Physics Letters B	*Phys Lett B*	136	385	211	259	–	991
Physical Review C	*Phys Rev C*	93	143	64	106	–	406
European Physical Journal C	*Eur Phys J C*	95	173	36	32	–	336
Astronomy and Astrophysics	*Astron Astrophys*	46	79	32	47	15	219
Nuclear Instruments and Methods in Physics Research A	*Nucl Instrum Meth A*	–	97	22	44	–	163
Physical Review B	*Phys Rev B*	53	96	–	–	–	149
Journal of Applied Physics	*J Appl Phys*	29	62	–	–	–	91
Astrophysical Journal	*Astrophys J*	29	–	–	27	–	56
The Lancet	*Lancet*	–	–	26	–	23	49
Nuclear Physics A	*Nucl Phys A*	–	45	–	–	–	45
Journal of Magnetism and Magnetic Materials	*J Magn Magn Mater*	–	44	–	–	–	44
Journal of the Korean Physical Society	*J Korean Phys Soc*	–	37	–	–	–	37
Monthly Notices of the Royal Astronomical Society	*Mon Not R Astron Soc*	–	–	22	–	15	37
Journal of Molecular Spectroscopy	*J Mol Spectrosc*	–	–	–	33	–	33
Journal of Dental Research	*J Dent Res*	–	–	–	27	–	27
International Journal of Systematic and Evolutionary Microbiology	*Int J Syst Evol Microbiol*	–	–	–	–	14	14
Antiviral Therapy	*Antivir Ther*	–	–	–	–	12	12
	Total	845	2338	951	1410	79	5623

Table 19.6 List of core journals of BRICS collaborative articles, indexed in the WoS, 2008–2015

Full name	Abbreviation	Brazil	Russia	India	China	South Africa	Total
Physical Review Letters	*Phys Rev Lett*	207	444	253	293	24	1221
Physical Review D	*Phys Rev D*	141	457	287	254	–	1139
Physics Letters B	*Phys Lett B*	169	243	188	207	26	833
Physical Review C	*Phys Rev C*	79	150	86	139	36	490
Journal of Applied Physics	*J Appl Phys*	38	65	42	41	–	186
Physical Review B	*Phys Rev B*	53	84	–	45	–	182
The Lancet	*Lancet*	–	–	44	37	40	121
European Physical Journal C	*Eur Phys J C*	41	77	–	–	–	118
PLOS One	*PLOS One*	–	–	42	46	28	116
Zootaxa	*Zootaxa*	–	51	–	54	–	105
Astrophysical Journal	*Astrophys J*	–	53	–	44	–	97
Acta Crystallographica Section E	*Acta Crystallogr E*	–	–	33	–	27	60
Journal of Clinical Oncology	*J Clin Oncol*	–	52	–	–	–	52
Astronomy and Astrophysics	*Astron Astrophys*	–	–	–	46	–	46
New England Journal of Medicine	*New Engl J Med*	30	–	–	–	–	30
	Total	758	1676	975	1206	181	4796

Again, the largest number of BRICS collaborative articles was published in journals classified in physics: three journals from this zone appeared in the lists for all BRICS countries: *Physical Review Letters*, *Physics Letters B* and *Physical Review C*. Although these journals appeared in the 2000–2007 list of core journals, this picture suggests a movement in BRICS, mainly in South Africa, towards an enlargement of the collaborative research among all members. Note that in the previous period, the five BRICS members published together in only one core journal, *Astronomy and Astrophysics*, published by the European Southern Observatory.

It is important to highlight that one of the most successful open-access journals appears in the list of core journals for three countries during this period—*PLOS One*, a journal classified in the field of multidisciplinary sciences. Similarly, *The Lancet*, one of the most prestigious journals in the medical sciences, appears in the list of core journals for three countries.

19.3.3 BRICS Collaborative Articles: Intellectual Structure

The data set presented in the previous section revealed that BRICS' collaborative articles tended to follow the general trends for BRICS' total publications (Table 19.3), with the field of physics in a central position. In order to better visualize the thematic trends, a journal co-citation analysis was performed for BRICS' collaborative papers. The co-citation analysis included only journals that were cited more than 100 times in each period. Figures 19.2 through 19.6 show each country's journal co-citation map.

For the Brazilian articles co-authored with other BRICS countries, the journal co-citation analysis generated a map with seven clusters for the period 2000–2007 (Fig. 19.2a). The largest cluster (violet) contains 15 (21.43%) of the 70 journals included in the analysis. The journals found in this cluster are predominantly related to biomedicine and general issues. *Nature* and *Science*, two major generalist journals included in cluster 1, have the highest degree of centrality. The second largest cluster (green) is strongly associated with physics and includes the journals with the largest numbers of citations: *Phys Rev Lett*, *Phys Rev D* and *Phys Lett B* (they can be identified on the map by the size of the nodes). The map presents five other communities, related to theoretical or applied physics (red, pink and dark red), astronomy/astrophysics (orange) and earth/-space sciences (light purple).

For the period 2008–2015, the Brazilian map exhibits eight clusters (Fig. 19.2b). Although the number of communities has increased compared to the previous period, the analysis shows a new arrangement for co-cited journals: the enlargement of the largest cluster (violet) and the reduction of some others, especially clusters 3 (light red) and 4 (pink). In this period, the largest cluster (violet) contains 48 (37.5%) of the 128 journals included in the analysis. The set of journals found in this cluster is mainly related to clinical medicine, and it does not include the journals with the highest degree of centrality, as was seen in the previous period. The profiles for clusters 4 (pink) and 7 (dark red) have also changed: cluster 4 is a multidisciplinary community, encompassing *Nature* and *Science*, which are again the journals with the highest degree of centrality, and cluster 7 is devoted to oral health. In ad-

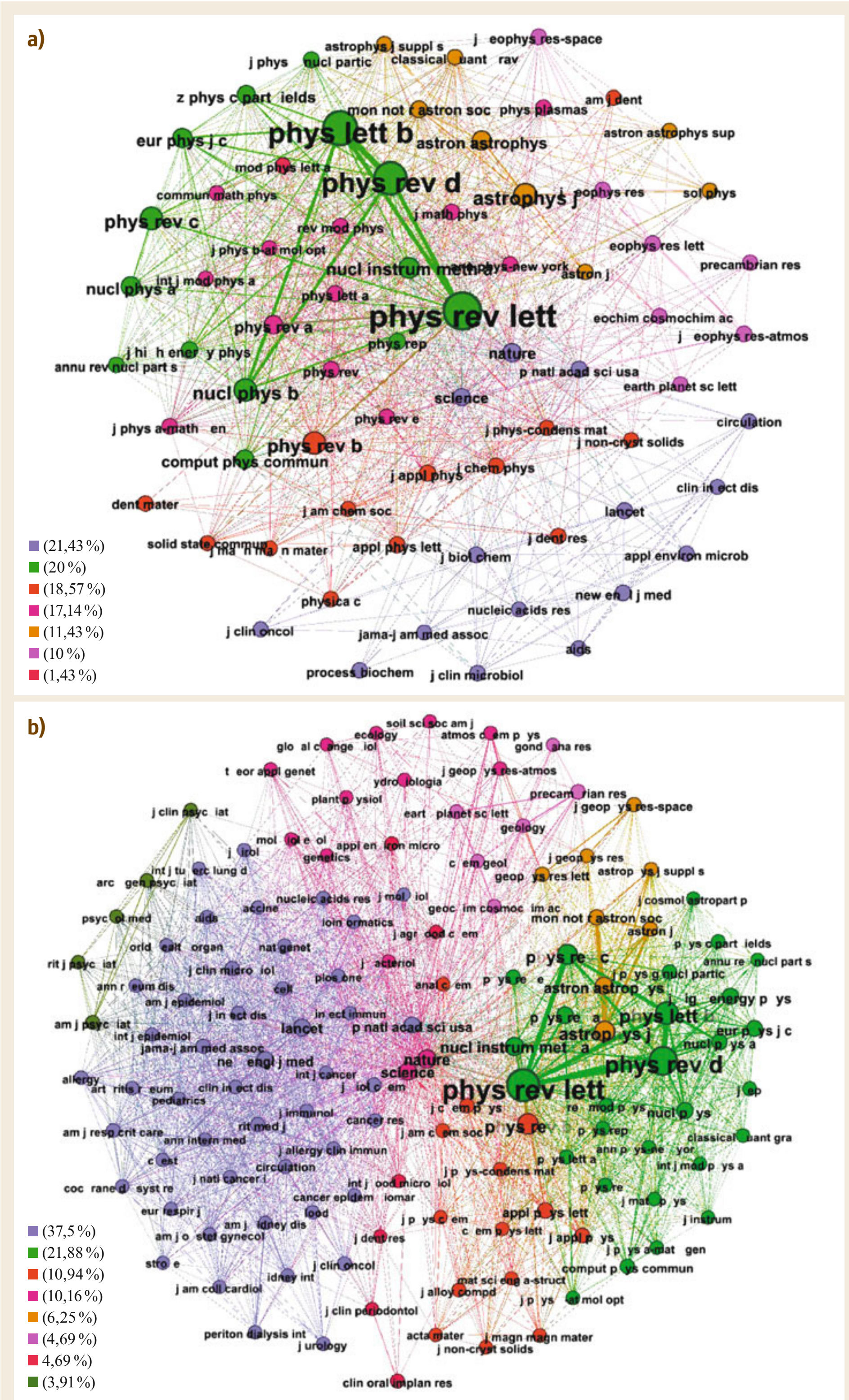


Fig. 19.2a,b Maps of co-cited journals of collaborative articles between Brazil and one or more other BRICS countries. Colors indicate different (thematic) clusters, while the size of the symbols represents the weight of the 70 **(a)** and 128 **(b)** journals in the analysis for 2000–2007 and 2008–2015, respectively. Gephi software (version 0.9.1)

dition, cluster 8 (dark green) has emerged in the field of psychiatry.

Despite these changes, some features remain unchanged from the previous period. Cluster 2 (light green), which is the largest, and cluster 3 (light red) are also related to physics. Cluster 2 still contains the journals with the largest number of citations, and they are the same as in the previous period: *Phys Rev Lett*, *Phys Rev D* and *Phys Lett B*. Clusters 5 (orange) and 6 (light purple) still comprise journals in astronomy/astrophysics and earth/space sciences, respectively.

Figure 19.3 displays the journal co-citation map based on Russian articles co-authored with one or more BRICS partners. The map for the period 2000–2007 presents nine clusters. The largest (violet) contains 45 (32.14%) of the 140 journals included in the map. Cluster 1 contains journals related to physics and chemistry. Cluster 2 (light green) is associated with physics, and it comprises one of the journals with the largest number of citations: *Phys Rev Lett*. Clusters 6 (light purple), 7 (dark red), 8 (dark green) and 9 (gray) are also associated with physics, while cluster 5 (orange) contains journals in astronomy and astrophysics (orange). Clusters 3 (light red) and 4 (pink) encompass journals in biomedicine and earth/space sciences, respectively. These clusters also include *Science* and *Nature*, journals with the highest degree of centrality.

With regard to the period 2008–2015, the map reveals only six clusters, indicating a compression of thematic communities. As observed for the previous period, cluster 1 (violet) remains associated with physics and chemistry. This is the largest cluster, with 64 (almost 35%) of the 183 journals included in this analysis. The other clusters display a new configuration compared to the previous period. Physics journals are found in clusters 3 (red) and 5 (orange), which include *Phys Rev Lett* and *Phys Rev Lett D*, the journals with the largest number of citations in this analysis. Clusters 4 (pink) and 6 (light purple) are now related to earth/space sciences and astronomy/astrophysics, respectively. Cluster 2 (green), the second largest community, encompasses journals related to biomedicine, including *Science* and *Nature*, which are again the journals with the highest degree of centrality in this map.

Considering Indian articles co-authored with other BRICS, the journal co-citation analysis generated a map with eight clusters for the period 2000–2007 (Fig. 19.4). Cluster 1 (violet) contains 15 (23.44%) of the 64 journals included in this map, which are mainly related to physics. Other small clusters are related to theoretical or nuclear physics: clusters 5 (orange), 6 (light purple) and 7 (dark red). These clusters contain the journals with the largest number of citations: *Phys Rev Lett*, *Phys Rev Lett B* and *Phys Rev Lett D*. Cluster 4 (pink) is related to astronomy/astrophysics, while cluster 8 (dark green) contains only one journal in mathematics (*J Math Anal Appl*). Finally, clusters 2 (light green) and 3 (light red) contain *Science* and *Nature*, the two journals with the highest degree of centrality in this analysis. Cluster 2 is strongly associated with the field of medicine, and cluster 3 with earth/space science.

For the period 2008–2015, the map based on journal co-citation analysis shows seven clusters, indicating a compression of thematic communities in India's collaborative papers. Cluster 1 (violet), which relates to physics and chemistry, contains 36 (23.68%) of the 152 journals included in this map. For this period, physics journals are observed in only one other community, cluster 4 (pink), which also contains journals with the largest number of citations: *Phys Rev Lett* and *Phys Rev Lett D*. Cluster 5 (orange) is related to earth and space sciences and contains *Science* and *Nature*, once again the journals with the highest degree of centrality in this analysis. Cluster 6 (light purple) comprises journals in astronomy and astrophysics, while clusters 2 (green) and 3 (light red) include journals in medicine and biomedicine, respectively. Cluster 7 (dark red) contains only one journal, *Ophthalmology*.

Figure 19.5 displays the map based on journal co-citation of Chinese articles co-authored with one or more BRICS partners in each periods. The analysis generates a map with eight clusters for the period 2000–2007. The largest cluster (violet) contains 33 journals (almost 28% of the total) related to biomedicine, with a strong emphasis in microbiology, virology, immunology and infectious diseases. Cluster 2 (light green) is mainly related to physics and chemistry, while clusters 3 (light red), 6 (light purple) and 7 (dark red) are related to nuclear and particle physics and theoretical physics. The journals with the largest number of citations are included in cluster 3 (*Phys Rev Lett* and *Phys Rev Lett B*) and cluster 6 (*Phys Rev D*). Cluster 4 (pink) comprises journals in earth and space science, including *Science* and *Nature*, which are the journals with the highest degree of centrality. Cluster 5 (orange) contains journals related to astronomy and astrophysics. The smallest community, cluster 8 (dark green), contains only one journal, related to veterinary science: *Small Ruminant Res*.

The map for the period 2008–2015 still holds eight clusters, but their profiles have changed substantially. Cluster 1 (violet) contains 72 journals (almost 28% of the total) and is related to physics and chemistry. Cluster 2 (light green) now contains multidisciplinary journals, including the fields of ecology, geophysics, genetics and biomedicine. This cluster also contains *Science* and *Nature*, again the journals with the highest

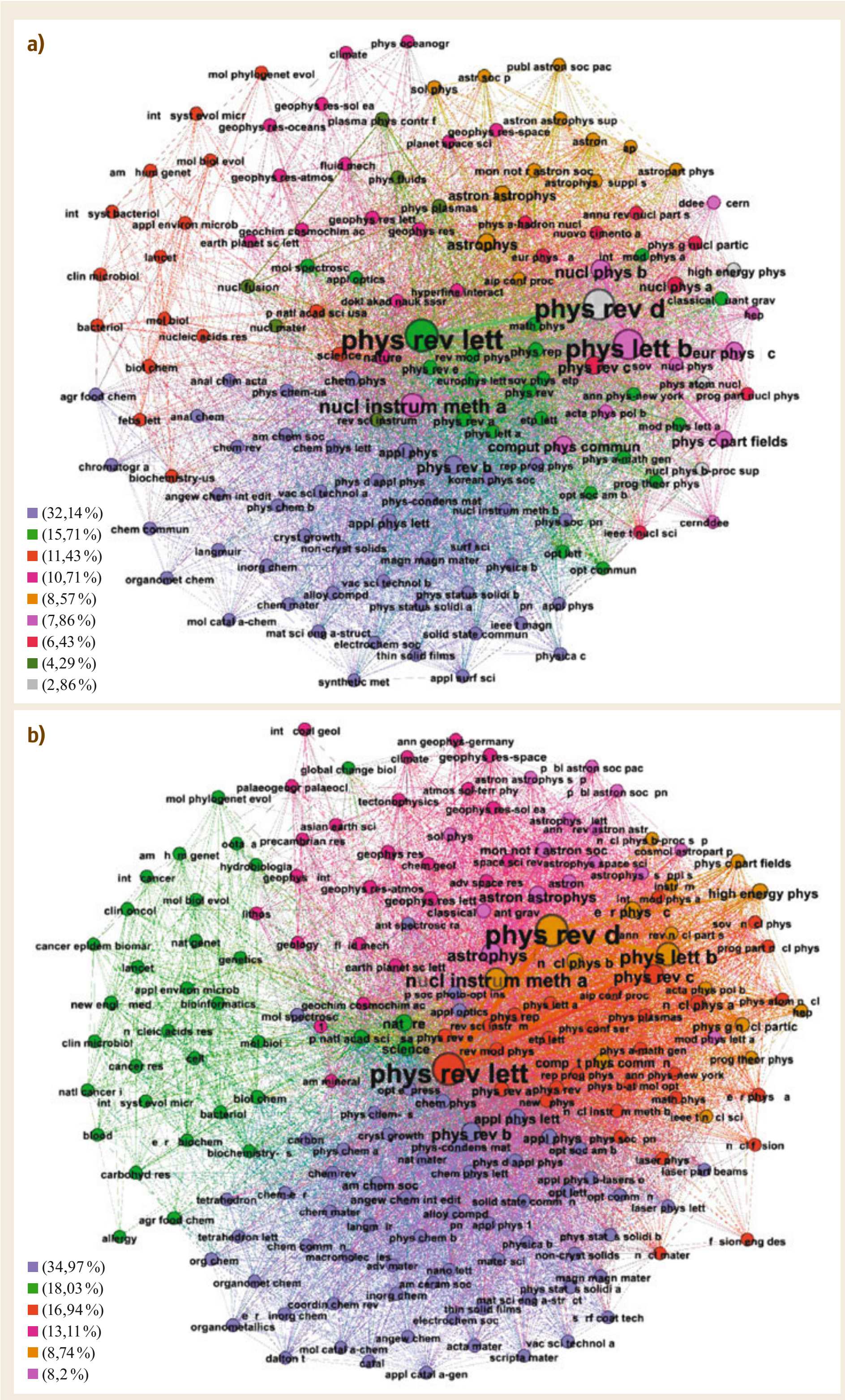


Fig. 19.3a,b Maps of co-cited journals of collaborative articles between Russia and one or more other BRICS countries. Colors indicate different (thematic) clusters, while the size of the symbols represents the weight of the 141 **(a)** and 183 **(b)** journals in the analysis for 2000–2007 and 2008–2015, respectively. Gephi software (version 0.9.1)

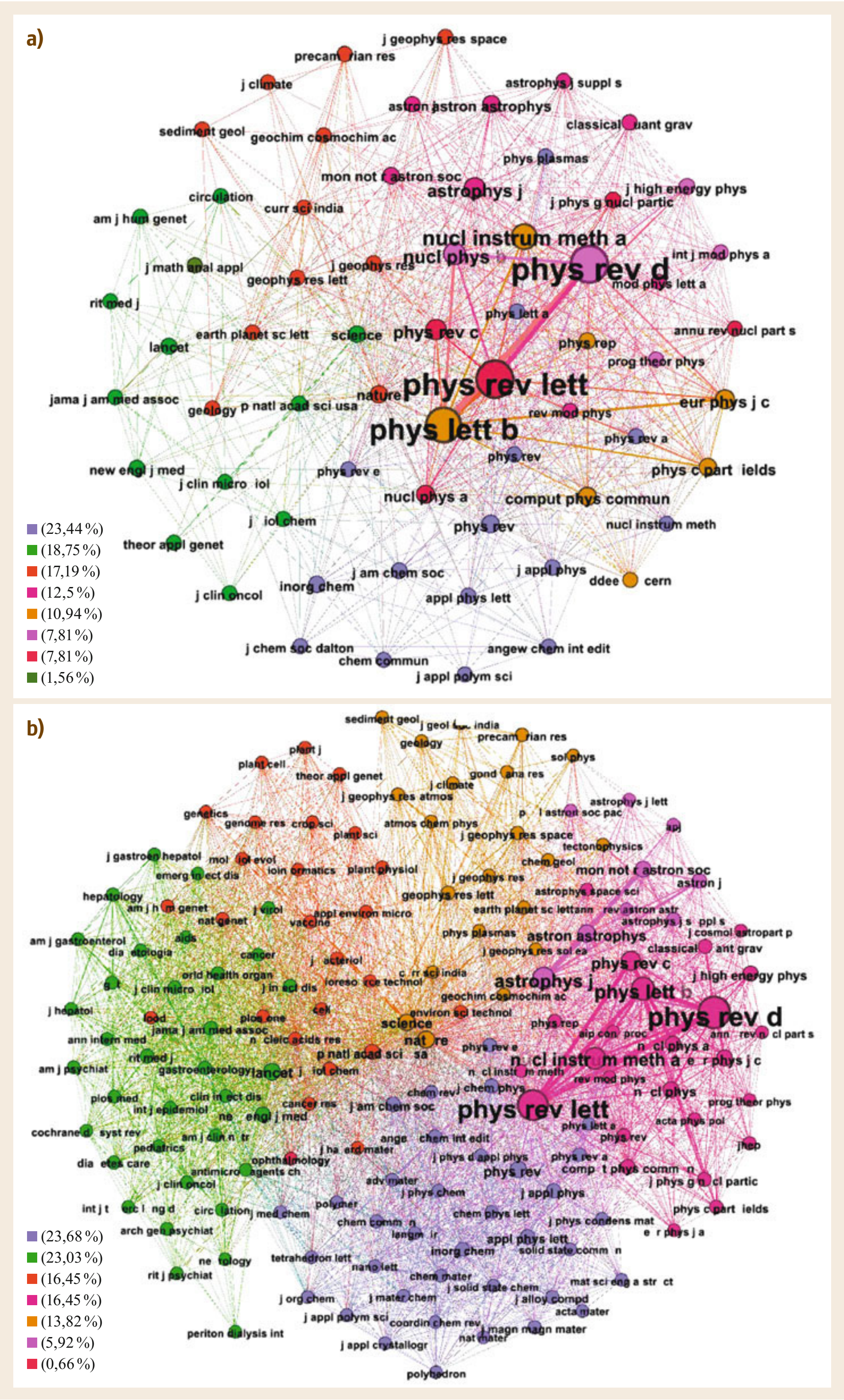


Fig. 19.4a,b Maps of co-cited journals of collaborative articles between India and one or more other BRICS countries. Colors indicate different (thematic) clusters, while the size of the symbols represents the weight of the 64 **(a)** and 152 **(b)** journals in the analysis for 2000–2007 and 2008–2015, respectively. Gephi software (version 0.9.1)

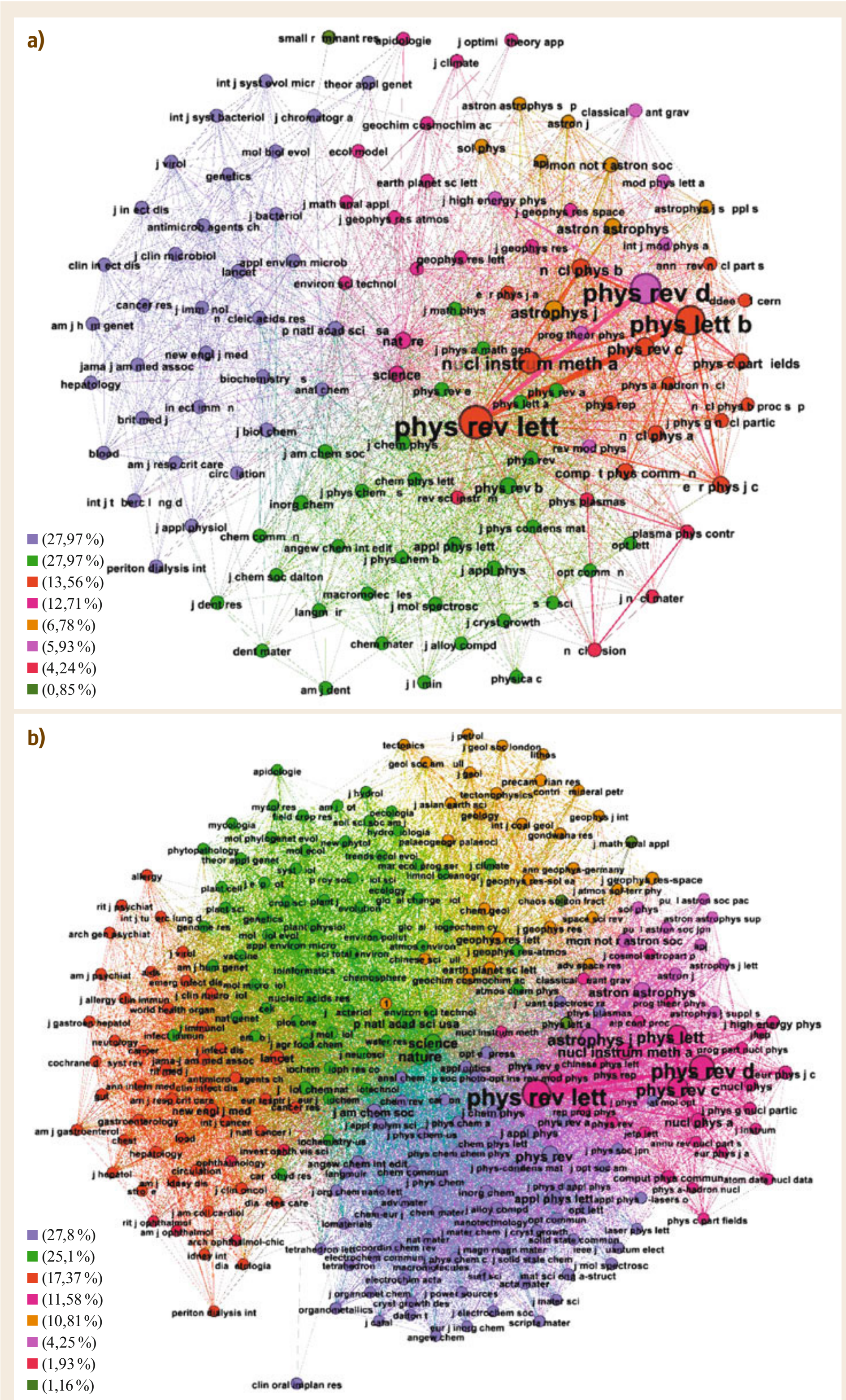


Fig. 19.5a,b Maps of co-cited journals of collaborative articles between China and one or more other BRICS countries. Colors indicate different (thematic) clusters, while the size of the symbols represents the weight of the 118 **(a)** and 259 **(b)** journals in the analysis for 2000–2007 and 2008–2015, respectively. Gephi software (version 0.9.1)

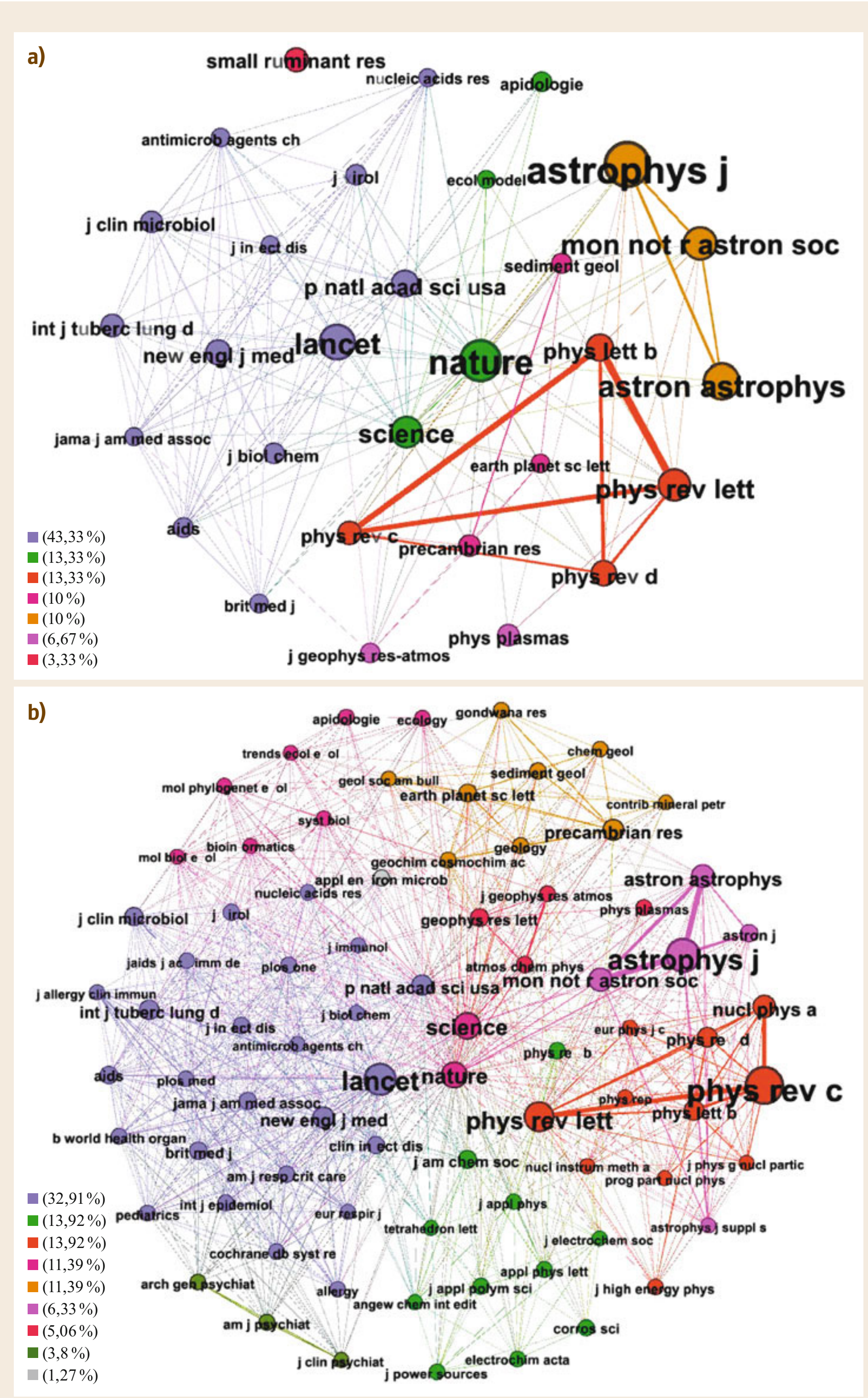


Fig. 19.6a,b Maps of co-cited journals of collaborative articles between South Africa and one or more other BRICS countries. Colors indicate different (thematic) clusters, while the size of the symbols represents the weight of the 30 **(a)** and 79 **(b)** journals in the analysis for 2000–2007 and 2008–2015, respectively. Gephi software (version 0.9.1)

degree of centrality. Clusters 3 (light red) and 7 (dark red) contain journals in clinical medicine and ophthalmology, respectively. Clusters 4 (pink), 6 (light purple) and 8 (dark green) comprise journals in physics (especially nuclear physics), astronomy and astrophysics, and mathematics and physics, respectively. The journals with the largest number of citations are included in cluster 4 (*Phys Rev Lett* and *Phys Rev Lett D*). Finally, cluster 5 (orange) comprises journals in earth and space sciences.

Finally, with regard to South African articles in collaboration with one or more BRICS countries, the analysis of journal co-citation generates a map with seven clusters for the period 2000–2007. As a first and general comment, South Africa's map is not as dense as those for the other BRICS countries, a result of the country's performance in terms of the total number of publications and number of collaborative publications (Tables 19.2 and 19.4). Cluster 1 (violet), the largest cluster, contains 13 journals in biomedicine. Other clusters have four or fewer journals, and thus are too sparse. Cluster 2 (green) contains *Nature* and *Science*, the journals with the highest degree of centrality. Clusters 3 (light red) and 6 (light purple) comprise journals in physics, and cluster 5 (orange) in astronomy and astrophysics. This cluster contains *Astrophys J*, the journal with the largest number of citations in this analysis. Cluster 4 (pink) encompasses journals in earth and space science; cluster 7 (dark red), with only one journal (*Small Ruminant Res*), is devoted to veterinary science.

For the period 2008–2015, the map displays nine clusters. The largest cluster (violet) contains 26 journals, still in the field of clinical medicine, with an emphasis in infectious diseases. Clusters 2 (light green), 3 (light red) and 7 (dark red) now contain journals in physics and chemistry, physics (mainly nuclear physics) and geophysics, respectively. Cluster 4 (pink) comprises journals in ecology and evolution, which are strongly linked to *Nature* and *Science*, those with the highest degree of centrality. Cluster 5 (orange) contains journals in earth sciences only, while cluster 6 (light purple) comprises journals in astronomy and astrophysics. Two small clusters, 8 (dark green) and 9 (gray), include three journals in psychiatry and one journal in applied microbiology, respectively. Unlike the maps for the other BRICS, here the journal with the largest number of citations includes not only those in physics, *Phys Rev C* and *Phys Rev Lett*, but also in astronomy and astrophysics (*Astrophys J*) and biomedicine (*Lancet*).

Table 19.7 summarizes various features found in the map analysis. The number of clusters per country varies from six to nine. Regarding changes in the number of clusters from one period to another, Brazil, China and South Africa show an increase, and Russia and India a reduction, indicating that as a whole, the thematic structure of BRICS collaborative papers did not increase after the group's formalization. Importantly, journals in the field of physics constitute those with the largest number of citations in both periods and across all countries. Also, *Nature* and *Science*, two prestigious multidisciplinary journals, appear with the highest degree of centrality in all countries for both periods. Although the main area of the largest cluster varied for some BRICS, the set of data presented in Table 19.7 indicates that there has been no relevant change at all.

Table 19.7 Map characteristics based on journal co-citation analysis of BRICS collaborative articles

Country	Clusters (n)	Centrality degree (mean)	Co-cited journals (n)	Largest cluster (area)	Journals with LNC[a]	Journals with HCD[b]
2000–2007						
Brazil	7	67	70	Biomedicine	*Phys Rev Lett*	*Nature*
Russia	9	139	140	Physics and chemistry	*Phys Rev Lett*	*Science*
India	8	63	64	Physics	*Phys Rev D*	*Nature*
China	8	111	118	Biomedicine	*Phys Rev Lett*	*Nature*
South Africa	7	28	30	Biomedicine	*Astrophys J*	*Nature*
2008–2015						
Brazil	8	123	128	Clinical med	*Phys Rev Lett*	*Nature*
Russia	6	180	183	Physics and chemistry	*Phys Rev D*	*Science*
India	7	148	152	Physics and chemistry	*Phys Rev D*	*Nature*
China	8	256	259	Physics and chemistry	*Phys Rev Lett*	*Nature*
South Africa	9	78	79	Clinical med	*Phys Rev C*	*Science*

[a] LNC = largest number of citations
[b] HCD = highest centrality degree
Astrophys J = Astrophysical Journal, Phys Rev Lett = Physical Review Letters, Phys Rev D= Physical Review D, Phys Rev C = Physical Review C

19.4 Discussion and Final Remarks

The present study investigated BRICS collaborative articles in order to map thematic trends before and after the formal establishment of the group in 2008. The primary purpose was to investigate whether the main themes in BRICS publications covered the priority areas identified in the Cape Town declaration. To this end, three main analyses were carried out: diachronic analysis of total publications, Bradford analysis and journal co-citation analysis of collaborative publications.

The results of the diachronic analysis indicated that, with the exception of Russia, the BRICS are also power countries in science, exhibiting a high annual increase in publication rates (Table 19.1). Such growth may be explained, at least in part, by their adherence to publishing in open-access journals (Table 19.2), which seems to boost the rate of BRICS publications. *Bouabid* et al. [19.39] also observed a strong growth rate for BRICS' scientific publications when analyzing their performance in 1995–1997 and 2010–2012, where the authors found that BRICS's growth surpassed that of the G7 countries.

With regard to the areas of research, this first set of analyses revealed that BRICS publications as a whole were largely oriented towards physics, chemistry and engineering (Table 19.3). *Yang* et al. [19.26] presented similar findings in an analysis of BRICS' disciplinary structure based on publications in 1991, 2000, 2009. *Kumar* and *Asheulova* [19.40] also found contributions in physics to be the most prolific among the emerging fields in Brazil, China and India. *Aksnes* et al. [19.41], who investigated the relative specialization index (RSI) over time, found that China's RSI performance differed from the global average, including some of BRICS members; the authors highlight that Chinese research is heavily concentrated in engineering and physical sciences. *Bouabid* et al. [19.39] found higher growth rates in total BRICS scientific publications not only in engineering and technology, but also in medical sciences, for two distinct periods, 1995–1997 and 2010–2012.

The analysis of BRICS' collaborative publications revealed two relevant features: collaboration varied widely among countries, and physics appeared as a central field in most of the BRICS countries.

The Bradford analysis based on BRICS' collaborative articles indicated a strong and major presence of physics journals among the set of core journals (Tables 19.5 and 19.6), corroborating results observed for total BRICS publications. However, a few journals, mainly related to medicine, were also included in the core journals, a finding that was not observed in the previous period.

The maps based on journal co-citation analysis confirmed the central role of physics in BRICS' collaborative publications. The maps indicated that BRICS scientific structures gained in density but not in diversity, revealing no relevant changes for the group as a whole since its formal establishment.

Comparing the two periods, the maps showed specific features by country (Figs. 19.2–19.6). The Brazilian maps indicated an important shift in the main theme from physics to medicine, including oral health and psychiatry. The Russian map showed a reduction in the number of clusters over the period 2008–2015, which was accompanied by an increase in the weight of physics. The Indian map showed a similar reduction in clusters, but physics was reduced as well. As for the Chinese, the maps revealed an increase in physics as well as the flourishing of new clusters related to medicine and multidisciplinary fields. Finally, South Africa exhibited an increase in thematic clusters, but the primary emphasis remained in biomedicine.

Journals in the field of physics exhibited the largest number of citations in both periods and for all BRICS countries. Physics also appeared as the main field in most of the clusters, as well as representing the largest cluster for three countries in the period 2008–2015 (Table 19.7). It is important to highlight that the group also demonstrated a steady potential for cooperation in other fields, especially in astronomy/astrophysics and earth-/space sciences, which was found in all BRICS' maps in the period 2008–2015. These findings indicate that BRICS' collaborative articles do have the intellectual underpinnings needed for cooperative efforts in energy, nanotechnology, space research and exploration, aeronautics, earth observation, geospatial technologies and climate change, which are considered priority areas in the Cape Town declaration.

Although an extensive share of the BRICS collaborative structure is in the field of physics, one may reasonably assume that such a strong level of cooperation probably involves other countries, and is not solely a reflection of internal BRICS collaboration. In fact, as observed in Tables 19.5–19.7, BRICS members collaborate in some fields of physics, such as particle physics and astronomy, which are well known from having large research teams from multiple institutions sharing high-tech facilities.

The constant presence of *Science* and *Nature* as the journals with the highest degree of centrality revealed some specificity for each country. During the period 2008–2015, these two most highly connected journals appeared together in various thematic clusters: earth and space sciences (Indian map, cluster 5), biomedicine

(Russian map, cluster 2), and ecology and evolution (South African map, cluster 4). These periodicals were also found in multidisciplinary clusters, which included a miscellany of journals, and represented the largest communities in the Brazilian (cluster 1) and Chinese maps (cluster 2). The fact that *Science* and *Nature* were connected and giving support to journals in different fields suggests an internal diversity among BRICS competencies, which is fundamental to driving cooperative efforts among the group.

Looking closely at internal diversity during the period 2008–2015, Russia's scientific structure exhibited a robust emphasis in biomedicine, while Brazil, India, China and South Africa were strong in clinical medicine. Hence, BRICS' collaborative articles displayed the intellectual foundation needed to develop research in medicine, biomedicine and life sciences, also included in the Cape Town declaration as priority areas. These findings are in accordance with the study by *Bai* et al. [19.22], which found that BRICS members were among the ten countries with the largest share of articles on neglected diseases.

The map analysis did not allow the identification of groups of journals related to some of the priority areas listed in the Cape Town declaration, including food security, sustainable agriculture, computing, water resources, information and communication technology, and science popularization. Nevertheless, the lack of identification cannot be seen as a complete absence of these issues in BRICS' collaborative articles. Certain factors must be considered in this picture, including methodological limitations. One such limitation involves the journal coverage in the international data source used for collecting BRICS' scientific publications, including collaborative articles. When compared with Ulrich's extensive periodical directory (around 63 000 journals), *Mongeon* and *Paul-Hus* [19.42] found that the journals covered by WoS represented 33, 28 and around 15% of journals in natural sciences and engineering, biomedical research and social sciences, and arts and humanities, respectively. The authors also found a strong overrepresentation of English-language journals. These findings confirm that WoS is not a comprehensive data source, especially for peripheral countries in science, such as most of the BRICS. Hence, a considerable number of BRICS scientific publications in national journals, mostly in local languages and in fields where research is more oriented towards local issues, were not included in the analysis. If such invisible literature were included, it could paint a different picture of BRICS' scientific performance, both generally and in terms of collaborative contributions.

Another methodological limitation is related to the analysis period. The groundwork is still being laid for the economic and political alliance between the BRICS countries. BRICS was formally established only in 2008, which may be a short period in which to observe strong interactions in science, especially if they are measured by a single variable such as scientific articles.

Despite these limitations, however, this chapter reveals interesting aspects of BRICS collaboration in science based on co-authored articles. Although no significant change in collaboration has occurred since the formalization of the group, the analysis suggests that the BRICS countries are working to enhance their cooperation in their main scientific competencies.

As collaboration in science is a complex phenomenon, effectively focusing on priority areas as the basis for enhancing BRICS' cooperation in S and T will demand a large effort among the whole group, including the establishment of common policies and initiatives to effect greater collaboration on the part of institutions and scientists among different fields within the BRICS. One remarkable initiative towards this end was the establishment in 2013 of the BRICS Think Tanks Council, with the aim of enhancing cooperative research. Council members include the Institute for Applied Economic Research (Brazil), National Committee on BRICS Research (Russia), the Observer Research Foundation (India), the China Center for Contemporary World Studies (China) and the Human Sciences Research Council (South Africa).

Policies aimed at changing the culture around mobility, a non-technical factor that promotes cooperation among countries, is as important as those mentioned above. A recent report on international student mobility shows that students from Brazil, Russia, India and China are highly attracted to OECD universities for study abroad, with 89% of Brazilian, 65% of Russian, 90% of Indian and 85% of Chinese students enrolled in an OECD university [19.43]. Such preference for mainstream countries has implications for future research ties among scientists, and represents an obstacle to enhancing and strengthening BRICS cooperation in science. Hence, the BRICS countries must develop strategies to promote a change in the culture of overvaluing cooperation with partners such as the USA, Canada, Japan and Western Europe. Otherwise, a true, effective partnership within the scientific community among BRICS countries may represent a mere utopia.

References

19.1 J. O'Neill: Building better global economic BRICs: Global economics, http://www.goldmansachs.com/our-thinking/archive/archive-pdfs/build-better-brics.pdf (2001)

19.2 M.R.S. Lima: Brasil e polos emergentes do poder mundial: Rússia, Índia, China e África do Sul. In: *O Brasil e os demais BRICs: comércio e política*, ed. by R. Baumann (IPEA CEPAL, Brasília 2010), http://repositorio.cepal.org/bitstream/handle/11362/1404/S3821B823D2010_pt.pdf?sequence=1&isAllowed=y

19.3 F. Mello: *BRICS e Cooperação para o Desenvolvimento Internacional* (IESC, Brasília 2015), http://www.inesc.org.br/biblioteca/publicacoes/livros/2015/brics-e-cooperacao-para-o-desenvolvimento-internacional/view

19.4 J.E. Cassiolato, H.M.M. Lastres: Science, technology and innovation policies in the BRICS countries: An introduction. In: *BRICS and Development Alternatives: Innovation Systems and Policies*, ed. by J.E. Cassiolato, V. Vitorino (Anthem, London 2009) pp. 1–34

19.5 BRICS Ministry of External Relations: Third summit: Sanya declaration and action plan, http://brics.itamaraty.gov.br/category-english/21-documents/67-third-summit (2011)

19.6 BRICS Ministry of External Relations: First BRICS science, technology and innovation ministerial meeting: Cape Town declaration, http://brics.itamaraty.gov.br/ptbr/category-english/21-documents/187-first-brics-science-technology-and-innovation-ministerial-meeting-cape-town-declaration (2014)

19.7 New York Academy of Sciences: Frontiers of Science, http://www.nyas.org/programs/frontiers-of-science/

19.8 Science News Staff: A look ahead at 2016: What research trends will be hot—and what will not?, http://www.sciencemag.org/news/2015/12/look-ahead-2016-what-research-trends-will-be-hot-and-what-will-not (2015)

19.9 OECD: *Science, Technology and Innovation Outlook* (OECD, Paris 2016), https://doi.org/10.1787/sti_in_outlook-2016-en

19.10 M.A. Bagchi: Comparative analysis of the factors for fostering innovation in BRICS countries from 1995 to 2009, J. Manag. **41**(1), 1–20 (2011)

19.11 A. Calderón-Martïnez, E. Ruiz-Conde: Leading emerging markets: Capturing and diffusing scientific knowledge through research-oriented repositories, Scientometrics **104**(3), 907–930 (2015)

19.12 E. Mauleon, D. De Filippo: Are the BRIC and MITS countries improving their presence in the international science? In: *Proc. 14th Int. Soc. Scientometr. Informetr. Conf*, Vol. 2 (Facultas, Vienna 2013) pp. 1868–1870

19.13 Y. Wang, J. Li-Ying: How do the BRIC countries play their roles in the global innovation arena? A study based on USPTO patents during 1990–2009, Scientometrics **98**(2), 1065–1083 (2014)

19.14 C.Y. Wong, L. Wang: Trajectories of science and technology and their co-evolution in BRICS: Insights from publication and patent analysis, J. Informetrics **9**(1), 90–101 (2015)

19.15 Y. Yi, W. Qi, D. Wu: Are CIVETS the next BRICs? A comparative analysis from scientometrics perspective, Scientometrics **94**(2), 615–628 (2013)

19.16 C.S. Wagner, S.K. Wong: Unseen science? Representation of BRICs in Global Science, Scientometrics **90**(3), 1001–1013 (2012)

19.17 L. Bornmann, C. Wagner, L. Leydesdorff: BRICS countries and scientific excellence: A bibliometric analysis of most frequently cited papers, J. Assoc. Inf. Sci. Technol. (JASIST) **66**(7), 1507–1513 (2015)

19.18 M.R. Guevara, M. Mendoza: Publishing patterns in BRIC countries: A network analysis, Publications **4**(20), 1–14 (2016), http://www.mdpi.com/2304-6775/4/3/20

19.19 U. Finardi: Scientific collaboration between BRICS countries, Scientometrics **102**(8), 1139–1166 (2015)

19.20 M. Singh, N. Hasan: Trend in research output and collaboration pattern among BRICS countries: A scientometric study. In: *4th Int. Symp. Emerg. Trends Technol. Libr. Inf. Serv.+* (2015), http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7048201

19.21 U. Finardi, A. Buratti: Scientific collaboration framework of BRICS countries: An analysis of international coauthorship, Scientometrics **109**(1), 433–446 (2016)

19.22 J. Bai, W. Li, W.M. Huang, Y. Guo: Bibliometric study of research and development for neglected diseases in the BRICS, Infect. Dis. Poverty **5**(1), 1–10 (2016)

19.23 R.N. Machado, J. Leta: Trends of intellectual and cognitive structures of stem cell research: A study of Brazilian scientific. In: *Proc. 14th Int. Soc. Scientometr. Informetr. Conf.*, Vol. 2 (Facultas, Vienna 2013) pp. 1759–1771

19.24 R.N. Machado, B. Vargas-Quesada, J. Leta: Intellectual structure in stem cell research: Exploring Brazilian scientific articles from 2001 to 2010, Scientometrics **106**(2), 525–537 (2015)

19.25 R.N. Machado, J. Leta: Crescimento, autores e temáticas da produção na área de células-tronco: o caso dos BRICS. In: *XVII Encontro Naci. Pesqui. Ciênc. Informaç* (PPGCI/UFBA, Salvador 2016)

19.26 L.Y. Yang, T. Yue, J.L. Ding, T. Han: A comparison of disciplinary structure in science between the G7 and the BRIC countries by bibliometric methods, Scientometrics **93**(2), 497–516 (2012)

19.27 B.C. Brookes: Bradford's law and the bibliography of science, Nature **224**, 953–956 (1969)

19.28 K.W. McCain: Mapping economics through the journal literature: An experiment in journal cocitation analysis, J. Am. Soc. Inf. Sci. **42**(4), 290–296 (1991)

19.29 M. Bastian, S. Heymann, M. Jacomy, M. Gephi: An open source software for exploring and manipulating networks. In: *Int. AAAI Conf. Weblogs Soc. Me-*

dia (2009), https://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154/1009
19.30 V. Blondel, J. Guillaume, R. Lambiotte, E. Lefebvre: Fast unfolding of communities in large networks, J. Stat. Mech. **2**, 100–108 (2008)
19.31 J. Testa: A view from Web of Science: Journals, articles impact, Inf. Serv. Uses **36**(1/2), 99–104 (2016)
19.32 J. Leta: Brazilian growth in the mainstream science: The role of human resources and national journals, J. Scientometr. Res. **1**(1), 44–52 (2012)
19.33 Y. Wang, R. Hu, M. Liu: The geotemporal demographics of academic journals from 1950 to 2013 according to Ulrich's database, J. Informetr. **11**(3), 655–671 (2017)
19.34 S. Yan, R. Rousseau, S. Huang: Contributions of Chinese authors in PLOS ONE, J. Assoc. Inf. Sci. Technol. **67**(3), 543–549 (2016)
19.35 D. Nicholas, P. Huntington, I. Rowlands: Open access journal publishing: The views of some of the world's senior authors, J. Doc. **61**(4), 497–519 (2011)
19.36 S. Lawrence: Free online availability substantially increases a paper's impact, Nature **411**(521), 521–522 (2011)
19.37 B.C. Björk, D. Solomon: Open access versus subscription journals: A comparison of scientific impact, BMC Medicine (2012), https://doi.org/10.1186/1741-7015-10-73
19.38 Centers for Disease Control and Prevention: Global disease detection program: South Africa, https://www.cdc.gov/globalhealth/healthprotection/gdd/south-africa.html (2017)
19.39 H. Bouabid, A. Paul-Hus, V. Larivierè: Scientific collaboration and high-technology exchanges among BRICS and G-7 countries, Scientometrics **106**(3), 873–899 (2016)
19.40 N. Kumar, N. Asheulova: Comparative analysis of scientific output of BRIC countries, Ann. Libr. Inf. Stud. **58**(Sept.), 228–236 (2011)
19.41 D.W. Aksnes, T.N. van Leeuwen, G. Sivertsen: The effect of booming countries on changes in the relative specialization index (RSI) on country level, Scientometrics **101**(2), 1391–1401 (2014)
19.42 P. Mongeon, A. Paul-Hus: The journal coverage of web of science and scopus: A comparative analysis, Scientometrics **106**(1), 213–228 (2016)
19.43 OECD: How is international student mobility shaping up?, Educ. Indic. Focus **14**, 1–4 (2013), http://www.oecd.org/education/skills-beyond-school/EDIF%202013--N%C2%B014%20(eng)-Final.pdf

Jacqueline Leta

Programa de Educação, Gestão e Difusão / Instituto de Bioquímica Médica
Universidade Federal do Rio de Janeiro
Rio de Janeiro, Brazil
jleta@bioqmed.ufrj.br

Jacqueline Leta is Associate Professor at the Institute of Medical Biochemistry of the Federal University of Rio de Janeiro, Brazil, where she is Head of the Science Diffusion and Education Department. Since 1999, when she received a PhD in Science Diffusion and Education, she has researched the performance of Brazilian science, with an emphasis on scientific publications. In 2005, she worked as a Postdoctoral student at Leuven Catholic University.

Raymundo das Neves Machado

Departamento de Fundamentos e Processos Informacionais, Instituto de Ciência da Informação
Universidade Federal da Bahia
Salvador, Brazil
raymacha@ufba.br

Raymundo das Neves Machado is Associate Professor at the Information Science Institute of the Federal University of Bahia, Brazil. He received a PhD in Information Science in 2015. His main research interests are: the formation and development of library collections, methods of evaluation of printed and electronic collections, studies in bibliometrics/scientometrics focused on some specific fields and/or countries, such of BRICS.

Roberto Lovón

Graduate Program in Information Science/IBICT-UFRJ
Federal University of Rio the Janeiro
Rio de Janeiro, Brazil
rlovon@gmail.com

Roberto Lovón is currently a Postdoctoral Fellow at the Graduate Program in Information Science at the Federal University of Rio the Janeiro (UFRJ), in association with the Brazilian Institute of Information in Science and Technology (IBICT). He received a PhD in Information Science from UFRJ-IBICT in 2015. Roberto holds a Masters in Production Engineering, with a focus on science and technology policies, and a Bachelor in Economics.

20. The Relevance of National Journals from a Chinese Perspective

Zheng Ma

The process of of journal evaluation began in the 1930s when the famous British scholar S.C. Bradford published his study of geophysics and lubrication, which presented the empirical law now known as Bradford's law of scattering, as well as the concept of core area journals. The citation indicator system and citation analysis theory system were founded in the middle of the twentieth century, and now have extensive influence. In the 1960s, Garfield carried out a large-scale statistical analysis of citations in journal literature. Generally speaking, the journal evaluation system has been gradually improved over time, producing an evaluation result that meets the development needs of science and technology. As one of the countries producing important science and technology outputs, China has ranked second according to the statistics of the number of scientific articles in recent years. At the same time, China has over 5000 scholarly journals, however, only 4% of them have been indexed in Web of Science and 10% of them in Scopus. A similar situation is found in Russia, Japan, Korea, and other non-English-speaking countries. Therefore, China has a lot of research and practice in the field of journal evaluation with which to explore more applicable and effective ways of assessing and improving national academic journal development. We will review the development situation of scientific, technical and medical (STM) journals in China to understand the demand for a national journal evaluation system. According to the comparative study on international and national evaluation systems and indicators of academic journals in China, we can find the characteristics of national journal evaluation under a framework of their respective evaluation purposes, evaluation methods, key features, and evaluation criteria. We introduce two cases of China's STM journal research and evaluation work: the development of the boom index and its monitoring function, and the definition and application of comprehensive performance scores (CPSs) for Chinese scientific and technical journals. English-language science and technology journals in China are more similar to international journals but are developing along a particular path. Therefore we also introduce three other cases: statistics and analysis of English-language science and technology journals in China, the communication value of Chinese-published English-language academic journals according to citation analysis, and the atomic structure model for evaluating English-language scientific journals published in non-English countries.

Part C | 20

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_20

Scientific and technological research activities in the field of natural science and scholarly communication and publication activities all show significant trends of globalization. The driver of this development comes from two sources. The first is the academic language, English, which plays an increasingly important role. Although Yuasa's idea on the shift of the global science center seems to predict that the world's science research center will leave the United States sometime in the future, the international research community seems to rely more and more on English as the sole de facto language of communication. Second, following large-scale trends, the massive volume of output from academic publishing continues to increase. Furthermore, mergers between science and technology publishing enterprises have become the norm. In this context, the *national belonging* of academic journals is itself a controversial and widely discussed topic.

In English-speaking countries, most academic journals are published in English. These English-language journals constitute the world's most extensive academic peer services, and generally they cannot be strictly defined as national or international journals. At the same time, the publishing of non-English-language journals seems to be walking in the shadow of English-language journals. Of the former, especially those from non-English-speaking developing countries, very few journals are included in international indicator systems. These journals can be viewed as *typical* of national academic journals.

In view of globalization trends, national academic journals could claim their value from four aspects. The first is at the academic resource level. In some research fields, including those with unique geographical and geological characteristics, such as traditional medicine among others, national academic journals tend to cover the most important academic research faster and more comprehensively and systematically. Second, at the methodology level, national journals might have unique and rational arrangements in terms of information organization and information architecture. For example, international journals usually provide research funding information in an acknowledgment section at the end the article. Most Chinese journals include funding information as a standalone annotation in the manuscript. Thus, the article funding ratio is a key indicator of a journal's quality. Inspired by this concept, the Science Citation Index (SCI) and other international search systems are beginning to include additional funding annotations. Third, regarding the efficiency of scholarly communication and behavior analysis, a huge number of national journals are an integral part of a country's scientific and technical output. China has over 5000 scholarly journals published in Chinese, 20 times that of English-language journals. A similar situation is found in Russia, Japan, Korea, and other non-English-speaking countries. Scientists, at the beginning of their careers, often publish important papers in national journals, which can be seen and confirmed by looking at the profiles of a number of Nobel Prize winners from China and Japan. Finally, in many countries governments provide substantial support for research and development activities, and therefore many achievements result from government-funded research and development. Thus, national researchers need to utilize this support as much as possible. Papers published in national journals are undoubtedly a rapid and effective way to achieve this. Furthermore, the readers and authors of national journals understand each other's scientific and social background, enabling the public supervision of academic ethics.

20.1 Journal Evaluation

20.1.1 The Origins of Journal Evaluation

Research on journal evaluation originates from philology [20.1]. Regarding the emergence of journal evaluation, scholars such as *Qian Ronggui*, *Qiu Junping*, and *Lai Maosheng* [20.2–4] noted that Western journal evaluation began in the 1930s. In 1934, the famous British scholar S.C. Bradford (b. 1878), in the study of geophysics and lubrication, presented the empirical law now known as Bradford's law of scattering. His research shows that

> if scientific journals in the size of the number of papers published in a certain subject, to reduce the order, then can be divided into a special subject for the core area of the discipline and the core of the same number of papers in several areas. At this time, the core area and successive districts of the number of journals into A1 A2 .... [20.5]

Based on this distribution, Bradford proposed the concept of *core area* journals, that is, many articles on a subject are published in the core area. The theory of foreign journal evaluation originates from this law.

Eugene Garfield was a famous American linguist. He helped to establish the Institute for Scientific Information and was one of the founders of scientometrics. The citation indicator system and citation analysis theory system were founded in the middle of the twentieth century, and have had extensive influence [20.6]. In the 1960s, *Garfield* carried out a large-scale statistical analysis of citations in journal literature, with the conclusion that many citations were concentrated in just a few journals and that a small number of citations were disseminated in many journals [20.7]. Subsequently, Garfield created the Institute for Scientific Information (ISI), and successively published the SCI, Social Science Citation Index (SSCI) and the Art and Humanities Citation Index databases.

Following the development of Western journal evaluation, efforts in journal evaluation are now emerging in China. In 1964, a number of Chinese scholars began to translate and use the SCI [20.8]. Gaining the attention of scholars, the theory of bibliometrics was soon introduced and disseminated in the 1970s. In the early 1980s, Chinese scholars translated and introduced several bibliometric laws. This included Bradford's law (1934) and Garfield's Law of Concentration (1955). These were followed by the literature aging index and citation peak theory in 1971 [20.2, 9]. These three laws are considered to be the theoretical basis for the quantitative evaluation of journals.

China's journal evaluation system was gradually established beginning with the introduction of foreign journal evaluation theory. In 1987, the Institute of Scientific and Technical Information of China (ISTIC) (commissioned by the Ministry of Science and Technology—formerly the National Scientific and Technological Commission) began to analyze the number of citations of Chinese scientific and technical researchers at home and abroad, and used statistical data to establish the Chinese Scientific and Technical Papers and Citations Database (CSTPCD) [20.10].

Based on research in the literature, this paper reviews the development of journal evaluation, the evolution of Chinese journal evaluation indexes from single indexes to composite indexes, and the change from object-oriented journal evaluation. The theory of journal evaluation has been developed over time, as has the establishment and development of various evaluation systems.

20.1.2 Development Trends Within the Evaluation System

An evaluation index typically involves two key methods of journal evaluation—a direct comparison according to the index of journal literature, and the evaluation of a composite evaluation index. The determination of core journals is ascertained from the initial single-index method to the production of a comprehensive composite index method. Furthermore, the journal evaluation system is gradually improved over time, producing an evaluation result that meets the development needs of science and technology.

When journal evaluation first began, a single index was used to determine core journals. With the development of evaluation work and the appearance of the composite index, evaluation work has become more scientific and reliable. According to the deviation of the function of each single index, the index can be divided into different categories such as evaluating the quality of journal papers or evaluating the performance of the journals [20.11].

The most common indicator for evaluating the quality of journal papers is Garfield's impact factor. Other indicators based on citation analysis include a journal's total citations, the rate of citation in other journals, and the percentage of deviations (as used by SCI and China's Citation Report for Chinese STM Journals) [20.12]. They also include the method of citation analysis of core journals proposed by *Garfield* [20.13], and the statistical method of the evaluation of core journals by means of important documents or indexes containing the number of journal papers [20.14]. In the evaluation of the volume of journals, this chapter proposes a method to determine Bradford's law of core journals. In general, these methods are used to measure the influence of academic journals as an alternative to the academic quality of academic journals. These methods include the annual index, the range of citations, citation rate, whether the full text of the article is reproduced in the journal, and the download rate. The index of the academic content in evaluating the ability of published journals includes the regional distribution of authors, the ratio of funded papers, and the number of citations [20.15].

Using a composite index to evaluate STM journals avoids narrow perspectives and the limitations of single-index evaluation. Furthermore, a composite index can comprehensively and objectively reflect the quality level of journals and is more suitable for the full evaluation of journals. Thus, this method has been widely used. Based on the principle of structural equations, an analytic framework of journal influence was devised by *Yue* and *Wilson* [20.16]. Previous studies state that the evaluation of Chinese humanities and social sciences journals is typically achieved using the indicator system [20.17]. Furthermore, a three-dimensional hierarchical structure of the journal evaluation indicator system was proposed, and the grey relational method was used to evaluate that system [20.18]. Others have used an analytic hierarchy process (AHP) to evaluate journals [20.19]. In addition, some scholars combine two or more evaluation methods. For example, the composite method is used to evaluate weights according to the AHP, and the weighted technique for order of preference by similarity to ideal solution (TOPSIS) method is one such evaluation method. Although the composite evaluation method has only one evaluation result, it can still be regarded as a multiple index comprehensive evaluation method in essence. The multi-index comprehensive evaluation of academic journals is a complicated system that involves many aspects including evaluation principle, index selection, data standardization, and evaluation method selection. Data standardization is used as a basic method of statistics. It should be no surprise that there have been few in-depth studies on the matter. It is necessary to point out that different methods of data standardization have had a considerable influence on evaluation results, and this chapter proposes a new standardization method of reverse index data based on the features of several common data standardization methods. A number of methods have been developed: *Luo Shisheng* prosed the comprehensive appraisal method [20.20], Suk the weighted synthesis method and the fuzzy linear weighted transformation method, and *Qin Lifu* looked at journal cost [20.21]. This chapter puts forward an appropriate selection method for core journals. Furthermore, Ma Wei devised the fuzzy comprehensive determination method, which looks at journal information density, intelligence reliability, and intelligence reporting speed. Lastly, *Wang Genbin* prosed the principal component analysis method from a mathematical perspective [20.22].

Composite indexes have been applied in the current evaluation system, and the citation reports of Chinese STM journals published by the China Institute of Science and Technology Information have been adopted as comprehensive evaluation scores. That is, the comprehensive evaluation indicator system used for Chinese STM journals includes the following features: it calculates many scientific measurement indexes, uses AHP to determine key index weights, evaluates each journal, and calculates the total score of each journal. This current evaluation method is very effective.

China's current journal evaluation system includes the following indexes: ISTIC's *Statistical Source of Chinese Science and Technology Journal* (also known as *China's Science and Technology Core Journals*), which publishes the Chinese STM Citation Report (CJCR); the Chinese Science Citation Database (CSCD), a source journal list issued by Chinese Academy of Sciences (CAS); Peking University Library's *A Guide to the Core Journals of China*; Chinese Social Sciences Citation Index (CSSCI) from China Social Sciences Research Center of Nanjing University, which covering the core *Journals of China's Humanities and Social Sciences* [20.23].

Bradford's aim was to distribute research, and this could be achieved via core journals. Journal readers can find references in journal articles and this will help the reader to meet their own needs more efficiently. Advances in science and technology and the rapid growth of scientific research have meant a change in the evaluation of journals. Initially, such evaluations were reader-oriented, and later there was a greater focus on libraries and other organizations. Because of a shortage of library funding in China, foreign core journals were translated into Chinese and introduced to China to ensure a better use of available subsidies [20.24]. In 1992, Peking University Library created its first publication outlining China's core journals; it was intended for use by the staff of Peking University Library. The aim was also to reduce costs in establishing a wide choice of journals [20.25].

The standardization of the journal evaluation system seeks to reduce the gap between journals and core journals. Such efforts are made to aid the future development and quality of journals, and for guidance purposes. Authors can also better understand the requirements of different journals, so as to help them contribute to and improve the quality of their paper.

In addition, the publication and management of journals is also an important object of journal evaluation, and the evaluation of core journals is bound to have an important influence on the editorial direction of journal publications. To maintain the continuous development of academic journals, journal management staff carry out effective research on journal evaluation as well as supervision.

Through core journals, we can quickly understand the dynamics and development of each subject or field. The journal evaluation system is regarded as a reference foundation, enabling libraries to optimize their collections. Librarians offer reading-list guidance and provide reference services, and the journal evaluation system can be used in such a manner. The system can also provide reference for the evaluation of academic achievements. Furthermore, it can provide reference for authors who wish to submit contributions and for government departments to manage journals. Core journals also promote the study of bibliometrics to provide information for journal development and competition, and to promote the development of journals and improve their quality. Therefore, core journals are not only important for the end users but also for the development of journals and their quality.

Indexes of core journals provide empirical evaluation measures, such as access law, expert investigation methods, and the reader survey method. With the development of bibliometrics, more and more methods based on mathematics and statistics are being applied to the evaluation of core journals, gradually enriching evaluation methods. Three main laws, Bradford's law, Lotka's Law, and Zipf's law, provide the theoretical basis for evaluation. They use bibliometrics to measure core journals according to certain basic steps [20.26]. Citation analysis is the most commonly used evaluation measure of academic journal quality. The corresponding quantitative evaluation indexes are as follows: impact factor, citation frequency, reaction rate, average citation rate, journal citation rate, and journal citation half-life. The most intuitive evaluation index is the journal impact factor, which is synthesized by the statistical analysis of citations and the citation phenomenon.

Bradford's law is the theoretical basis of the core journals. Garfield and others looked at citation frequency to evaluate science and technology journals. The advent of this method in academia marked a milestone in the evaluation of the journal. However, there are some limitations in using the total citation frequency to evaluate a journal. Garfield revised his own research and proposed the impact factor index to evaluate journals. The impact factor is more reliable than the total number of citations, as it eliminates the influence of many uncertain factors. The impact factor has become a universal journal evaluation index, and the main measurement index to evaluate the academic influence of journals.

Total Citations: The total number of papers published by a journal since its inception and published in the year in which it was cited. The total citation frequency is a very objective and practical evaluation index, and can be used to measure the academic influence of journals. It can also be used to objectively explain the degree of use and attention of journals, and the position and function of journals in academic exchange. It does face the problem, however, where the impact factor will be low if there is an excessive volume of journals. The higher the frequency of citations, then the higher the value of the journal's utilization and academic level [20.27]. However, citation frequency is also subject to the publication of the journal, the number of papers published, and whether the publication is a professional journal or a comprehensive one [20.28].

Impact Factor: The number of citations received in that year by articles published in that journal during the two preceding years, divided by the total number of articles published in that journal during the two preceding years [20.29]. It reflects the overall academic level of the journal, generally in the same subject area: the larger the impact factor, the greater the impact of the journal [20.30].

Three basic factors are important when calculating an impact factor: the number of papers, time, and the number of citations. Furthermore, the peak period of a paper is dependent on the period of time. A scientific paper's peak citation period is claimed to be two years after publication, according to scientific metrology, so the current international practice is to set a period of two years.

Reverse Indicators: The index of reverse evaluation in journal evaluation concerns articles that received zero citations. This can be traced back to the 1960s, when Price, one of the fathers of scientific metrology, first looked at basic statistics regarding papers that received zero citations published in *Science*. He found that approximately 35% of the papers in any statistical time window had not received any citations [20.31]. Regarding zero citations, Chinese scholar Wu Yishan has proposed that we not forget the *Sleeping Beauty* phenomenon. A lack of citations may not necessarily mean poor quality research results, and zero citations may represent an excellent result from another perspective [20.32].

20.2 Development of STM Journals in China and Demand for Evaluation

20.2.1 The Status of STM Journals in China

In recent years, the number of published journals and the number of STM journals has remained stable. According to statistics from a *National Press and Publication Situation* report, China published 9867 journals in 2012, of which about half were natural science, with technical journals accounting for 4953. Since 2007, the annual growth rate of total journals and STM journals in China has been less than 1%.

The ISTIC has selected about 2000 high-quality and influential academic and technical journals as Chinese STM core journals. The institute uses quantitative indices to monitor the development level and trends of Chinese STM journals, and publishes the annual CHCR. China's core journals in science and technology published close to 500 000 papers in 2014, with an average of 260 papers published per journal each year.

Regarding journal evaluation, total citation frequency and impact factor are the indicators typically used to measure the absolute influence and relative influence level of journals. In 2012, the average citation frequency of Chinese STM journals reached 1023 citations per issue, 4.5 times that of the 2001 rate, and the average impact factor was 0.493, 1.7 times that of 2001. The article funding ratio index refers to the proportion of the results of provincial- and ministerial-funded projects published in a journal. This index reflects the ability of the journal to attract high-level research papers. The average value of the article funding ratio in China's journals was 52% in 2012, which means that more than half of the papers in the core journals were the research results of various nationally funded projects. The 2001 value of the fund thesis was 34%. The significant growth of the above indicators shows that China's STM journals play a significant role in the dissemination of knowledge and academic exchange.

The publishing cycle of China's STM journals has been gradually reduced. In 2007, the monthly ratio of China's STM core journals was 28.73%, rising to 35.79% in 2011. There was also a considerable shift from bimonthly publications to monthly, with a quarterly total of 13.22% in 2008 compared to 10.66% in 2011. The data show that China's STM journals are increasing in publication speed, reducing the journal publishing cycle and improving the timeliness of knowledge dissemination. At the same time, the total number of papers in Chinese STM core journals remains at around 500 000 a year.

Although the number of STM papers published in China ranks second in the world, the academic quality and influence of STM journals remains low. The SCI database is the world's most widely used science and technology literature retrieval system. Impact factors and other classical indicators are used to choose more than 8600 high-level scientific and technological journals worldwide (mainly English-language journals). According to the ranking of impact factors in corresponding disciplines, SCI journals are divided into four main categories: the impact factors determine whether a journal is classified as Q1, which indicates that the journal is a leading publication in that particular discipline, followed by Q2, Q3, and Q4. In 2015, 152 Chinese journals were covered in SCI, accounting for less than 2% of all journals published in China. Twenty journals were considered to be at the fore of their particular subject area, and were thus classified as Q1 journals, representing just 8% of Chinese STM journals. A total of 40 journals were classified as Q2, just 15% of all Chinese journals.

The world's largest scientific and technological publishing organization, the Elsevier Scopus database, is an important multilingual document retrieval database. It uses SCImago Journal Rank (SJR) and other indicators to assess and collect more than 12 000 major global scientific and technical journals. The SJR index also ranks journals from Q1–Q4: a journal considered a Q1 journal in the subject category ranking indicates that the journal is a leading journal in that particular subject area, followed by Q2, Q3 and Q4 journals. According to Scopus data, there are 554 journals from various areas in China, representing less than 5% of all Scopus journals. It is very difficult for Chinese journals to achieve a Q1 ranking, with just 27 journals considered Q1 journals, accounting for 5% of China's Scopus journals, with 152 (27%) ranked as Q2 journals. Thus, the majority of journals do not achieve such positive rankings.

20.2.2 The Main Evaluation Methods of National Science and Technology Organizations

China's Ministry of Science and Technology

Our government ministries have always attached great importance to the quality management of academic journals. Furthermore, it is common for third-party organizations to evaluate such journals. The journal evaluation system, which has had a considerable influence on the evaluation of STM journals, is a comprehensive evaluation indicator system of Chinese STM journals developed by the China Science and Technology Information Research Institute (CITIC Institute). The CITIC Institute has been developing evaluation indexes for Chinese STM journals since 1987, and publishes an annual citation report, looking at three key levels: academic quality, international competitiveness, and sustainable development potential. To comprehensively reflect the quality and influence of academic journals through more than 20 academic indexes of STM journals, the evaluation results are widely used by the General Administration of Press and Publication, the Ministry of Science and Technology, the Ministry of Education, the National Natural Science Foundation Committee, the CAS, and other industry departments and institutions of academic journal evaluation. This system has achieved good results. Since 2000, the CITIC Institute has been commissioned by the Ministry of Science and Technology to conduct pioneering work in the research of high-level STM journals. The institute has helped to outline a strategy of high-quality Chinese STM journals, develop an evaluation indicator system of such journals, and construct a service and guarantee system for these publications. Furthermore, every three years it publishes a list of China's high-quality STM journals (starting in 2008). The CITIC Institute is considered impartial and authoritative. Aside from these evaluation activities, it also plays a guidance role for STM journals. Under the promotion of the evaluation system of CITIC's STM journals, the overall level of our country's STM journals has improved. Journal cataloging has been standardized and internationalized, and the quality of journals has improved correspondingly.

In 2012, the CITIC Institute launched the *Leader of the 5000—China's top-level academic paper platform* (F5000). The F5000 project is the continuation of the high-quality STM journal project that annually selects 5000 outstanding articles from the top journals to further promote the brand image of *Chinese quality STM journals*. It is also important to promote the improvement of the overall academic level of Chinese STM journals and quality scientific research in China. Further aims include improvements to academic communication and the dissemination of knowledge, and the development of our academic influence and international competitiveness.

Ministry of Education

In 2006, the Ministry of Education published the results of the first *High-quality Chinese University STM Journals* report. STM journals from universities and colleges were classified according to the professional attributes of the organizers and journals, and in accordance with the needs of the development strategy for national quality STM journals. It established a model for university STM journals, and identified the desired characteristics of such journals.

China Association for Science and Technology

With the aim of promoting the innovation and development of Chinese STM journals and enhancing the core competitiveness of STM journals, the China Association for Science and Technology aims to continue to promote quality STM journals in China, under a project to cultivate quality STM journals, the international promotion of STM journals, and the development of science impfaction. In this way, English-language STM journals in China will be further developed. Additionally, a national service innovation system will be created. This will work to strengthen China's position in the international science and technology arena, and our cultural soft power. The China Association for Science and Technology, the Ministry of Finance, CAS, the Ministry of Education, and the China Institute of Engineering will work together to achieve *The Plan to Promote the International Influence of China's Scientific and Technological Journals*. The funding for the project is 100 million yuan, which represents the full support of China's ministries.

Chinese Academy of Sciences

CAS established the Science Publishing Fund of the Chinese Academy of Sciences in the early 1990s to support quality and important scientific and technical publications, advance scientific and technological publications, and promote the development of science and technology. The scope of the funding is to help key science and technology journals in various fields of natural science and technical sciences. The focus of funding on national key projects reflects the results of many papers, domestic and foreign high-frequency citations, impact, and the quality and efficiency of high-level science and technology journals.

20.2.3 The Demand for a Journal Evaluation System for Journal Development

Is it possible to evaluate a reasonable result under the same evaluation system in different stages of development? For the evaluation of academic journals based on different stages of development, such as the current evaluation system for the selection of *core journals* or *source journals*, the comprehensive journals of colleges and universities and CAS social sciences journals are classified as separate *comprehensive categories*, in parallel with the major journal categories. The evaluation of these two major journal types in the comprehensive category adopts the same method as the evaluation of professional journals. That is, the various disciplines of a comprehensive journal are regarded as a discipline, and the data of the whole journal is accumulated. This evaluation method has played an important role in improving the influence of university journals and comprehensive social science journals in the initial development stages. However, because of the emphasis on each subject, the difference between citation frequency and the impact factor of different subjects can sometimes be substantial, making the evaluation problematic [20.33]. For example, the following five categories are applied to China's STM journals: policy journal, academic journal, application journal, abstract journal, and popular science journal. These five categories of journal are all very different to each other, and so the assessment criteria system for each category of journal should be unique from all others [20.34]. In the evaluation of different types of journals, only one quantitative index or influence factor is used, and this practice has proved unsuitable.

SCI, founded in 1964 by the Institute for Technology Information, is one of the six largest retrieval systems in the world. In recent years, China has been attaching greater importance to SCI, which is often regarded as the main index to evaluate the level of scientific research of a unit and an individual. SCI includes journals with a high impact factor and a high level of scientific research. SCI can be used to evaluate the level of scientific research indicators, and can well reflect the scientific level of researchers. At the same time, low-level staff, if seeking the evaluation of scientific research, use SCI to evaluate the level of scientific research indicators. There will be a clear evaluation of what is inappropriate and not applicable.

Core journals are widely applied to the performance evaluation of scientific research. Although these have a certain evaluation function in scientific research work, their present function in research performance evaluation is exaggerated. It is not scientific to substitute the quality of a paper for the journal quality in the evaluation process. The aim of the core journals is to optimize the collections of library and information departments and to provide a reading service for readers. This has a certain objective reference effect on the performance evaluation of scientific research. When the scientific research management department carries on the scientific research performance evaluation, whether the scientific research results are published in the core journals, is often a key reference factor. However, other excellent scientific and comprehensive journals exist outside the core journals. Usually core journals are selected and other outstanding journals are ignored. Thus, the core journals should not be used to evaluate the performance of scientific research.

The existing evaluation criteria and evaluation indicator system (usually from journals, citations, and third-party evaluations of three dimensions to build an evaluation indicator system) along with people's understanding of the evaluation work has been deepened. The evaluation method has made great progress; each evaluation institution has adjusted the direction of their evaluation and has optimized its evaluation index. Evaluation in China has become even more refined. Furthermore, it has expanded from two or three initial indexes to nearly 20, although the evaluation indicator system adopted in each published evaluation report has changed or expanded from the original foundation. Some problems, however, have not been solved, including political quality, academic quality, editorial quality, and publication quality. These are important quality standards of journals, but the current evaluation of *core journals* based on impact does not directly embody political quality, and the evaluation indexes related to journal quality (e. g., editorial, institution, printing, and binding) are not considered [20.35]. As an important evaluation index in the evaluation system, impact factors (because of abnormal interactions among journals) suffer human interference, and the effect of download rate, click rate, and the academic ecological environment is destroyed. Therefore, the evaluation of academic journals faces great difficulties and problems. It is then necessary for the evaluation indicator system to be constantly adjusted and improved.

The evaluation of academic journals in China is currently facing a number of challenges such as the time lag in journal evaluation and the opacity of the evaluation process. When the evaluation results of core journals are applied to the evaluation of scientific research, many problems arise. The name-publication effect of the evaluation system has greatly tightened the environment of noncore journals and has essentially destroyed the academic environment. The premise of scientific evaluation results is the existence of a healthy academic environment. Once this environment is damaged, the good evaluation index also becomes flawed. Thus, China's system of journal evaluation has many problems including the situation concerning systematic evaluation results and the Matthew effect of core journals, which influences academic environment, and so on.

20.3 Comparative Study of International and National Evaluation Systems of Academic Journals in China

20.3.1 Overview of the Major International Journal Evaluation Systems

The international journal retrieval and evaluation system was developed and continues to be developed following the concept of Bradford's core journals and the development of bibliometrics. The process of journal evaluation is a standard measure of academic quality and the overall level of journals, with one aim being the reduction of disputes. At present, the widely recognized journal evaluation system comprises SCI and the Engineering index (Ei) published by Elsevier.

SCI, founded in the 1960s, is a citation database that was originally used for retrieval and was then developed to determine the international influence by journals. The Journal Citation Reports (JCR) established in 1975, are a tool for the quantitative evaluation of SCI-indexed academic journals. Because of the influence of SCI, the journals or papers included are generally considered to have higher academic level and greater influence.

Ei was established in 1884, and mainly contains engineering and technical journals, conference papers, and science and technology reports. As a widely recognized abstract retrieval tool, Ei has considerable authority in the field of engineering technology.

Scopus is the world's largest digest and citation database, developed by Elsevier. It concerns the fields of science and technology, medicine, and social sciences in peer-reviewed academic literature and high-quality network resources. Scopus also provides visual intelligence tools, aimed at providing tracking and analysis of research service results.

The MEDLINE database covers biomedical and life sciences topics vital to biomedical practitioners, educators, and researchers. This database covers publications from 1966. The literature represents the highest level of global medicine and provides the latest biomedical research results and trends.

J-STAGE is an academic journal network platform founded by Japan's Science and Technology Agency in the late twentieth century. It mainly includes Japanese science and technology journals in fields such as physics and computing.

SCI

SCI is a citation database launched in 1964 by the ISI. The SCI CD-ROM database was established in May 1988. Since 1975, ISI has published the annual worldwide JCR based on SCI. JCR is an effective tool for the quantitative evaluation of academic journals. SCI has a wide range of disciplines, academic influences, and extensive coverage. Thus, the journals and papers included in SCI reflect a high academic level and significant international influence. SCI is the most powerful tool in the evaluation of scientific research, science and technology, scientific research institutions, science publications, and science subjects themselves. SCI is the authoritative retrieval system of scientific and comprehensive engineering papers from all over the world. In addition to scientific evaluation, SCI's unique role has been widely recognized worldwide. Garfield stated: "A valid index must strictly limit the scope of its inclusion, only to collect information useful to the researcher." Therefore, we should look at and use the authoritative citation data of SCI.

Main Features. Looking at the external features, the main features of SCI are obvious, and include science and technology journals in the field of natural science as well as many journals from many countries. That is, it has a wide international scope, with an impressive and comprehensive number of journals covering many decades. Furthermore, it has a relatively short renewal cycle, resulting in excellent information timeliness.

Producer. ISI merged with Thomson Reuters in 1992. It is currently owned by Clarivate Analytics.

Target Journals. SCI covers all natural science disciplines, with a total of 273 categories. It is an improtant Journal Abstracts Index as it covers the most important journals in almost all science field specially in those subjects of basic theory of natural science. It can be used to retrieve journals in the following areas: mathematics, physics, chemistry, astronomy, biology, medicine, mechanics, optical engineering, science and technology, power engineering and thermal physics, electrical engineering, electronic science and technology, computer science and technology, civil engineering, aerospace science and technology, control science and engineering, environmental science and engineering, food science and engineering, management and education, materials science and engineering, and information and communication engineering. It has been providing important academic achievement information since 1945.

Purpose of Evaluation. It is convenient for researchers to conduct a comprehensive literature search to understand the historical evolution, the influence, and the development trends with the subject area. Furthermore, scientific research management departments can use SCI statistical analysis data as a quantitative basis for the performance evaluation of scientific research institutions and personnel.

Evaluation Method. Evaluation moves from qualitative evaluation to quantitative evaluation. Every year Chinese journals are audited, with approximately 200 selected. The selected journals will also be audited to ensure that they maintain their high level of quality. Each journal is subject to a broad evaluation process before being selected or eliminated. The editors responsible for the evaluation work have an appropriate educational background and professional experience and training in the relevant fields. Experts from various fields may be brought in where necessary.

Selection Criteria. From the perspective of requirements and procedure, SCI focuses on the quality of journal content editing, academic quality, and academic integrity. It also pays attention to the international and academic level of authors and editorial boards, and emphasizes sustainable development.

Evaluation Index. SCI mainly uses impact factor and citation frequency.

According to the upper introduction and study materials about SCI, we separat those raw materials into deconstruction items and re-organize them as reconstruction item as similar structure in order to comparative study different evaluation system. See Tables 20.1 and 20.2.

Ei

Ei (now Ei Compendex) was first published in 1884 by the United States Engineering Information Corporation, which publishes journals, conference papers, and scientific reports on engineering and technology disciplines. The literature contained in the database covers almost every field of applied engineering technology. Ei is a comprehensive and abstract retrieval tool in the field of engineering technology. It does not generally report on science literature focusing on pure theory, nor does it pursue a massive collection. Instead, attention is paid to the quality of the literature. Ei Compendex is the world's oldest database of engineering abstracts.

Main Features. Ei includes literature covering all engineering fields. The database contains 3639 of the world's leading engineering journals from more than 70 countries. It has been a respected provider since 1884, and 90% of the literature is in English. Ei retrieves weekly updates.

Producer. Elsevier.

Target Journals. Ei concerns engineering technology and covers more than 190 disciplines such as power, electrical, electronics, automatic control, mining and metallurgy, metal technology, machinery manufacturing, management, civil engineering, water conservancy, and education projects. The Ei retrieval system has a high level of comprehensiveness, a wide data source, extensive geographical coverage, wide coverage, high quality, and strong authority.

Purpose of Evaluation. To provide professional and practical online data and information services for scientific researchers and engineering technicians.

Selection Criteria. The Ei profile program consists of five main areas:

A. The primary criterion is journal subject, with Ei largely focusing on chemistry, computer engineering, and software. However, journals in the fields of agriculture, industry, textiles, applied chemistry, mathematics, and atmospheric science are included. More general fields may also be included if it is a very important publication; the value of the article determines whether it is included within the scope of Ei. Excluded journals are those in the areas of biology and astronomy.
B. Ei journal information generally includes journal name, ISSN, fax, e-mail, and similar information.

Table 20.1 SCI features and requirements

Raw materials	Deconstruction	Reconstruction
(1) SCI focuses on the field of natural science	(1) The field of natural science	(1) Subject characteristics
(2) SCI contains more than 80 countries	(2) Digest index database	(2) Database properties
(3) SCI contains more than 12 000 world-leading STM periodicals	(3) Basic theory	(3) World influence
(4) SCI covers more than 250 disciplines in the field of key academic results since 1900	(4) Involving national journals (5) Covering disciplines	(4) Efficiency and timeliness
(5) SCI database update frequency is once a week	(6) STM periodicals	(5) Journal attributes
	(7) Update frequency	
	(8) Academic achievements	
	(9) Areas of focus	
	(10) Time range	

Table 20.2 SCI requirements and process

Raw materials	Deconstruction	Reconstruction
(1) SCI believes that periodical publication is one of the most important indexes to measure periodicals, and it is very important to publish periodicals on time according to the publication cycle. Timely publication allows readers to obtain the latest information and reduces the collection of published early or outdated periodicals. Before completing the periodical evaluation, the editors usually follow three contiguous issues of the publication. (2) Sci-indexed periodicals require journal editors to conform to international editorial conventions, such as full number of addresses (3) English titles, abstracts, headings, keywords, etc. (4) SCI will peer review as an evaluation index, and to some extent it can be considered that through peer review, journal paper quality is also guaranteed (5) The periodicals included in SCI need to publish moral statements and practices that do not to accept falsehoods and academic misconduct (6) The periodical publication format printing or electronic format (XML) is qualified to participate in the evaluation (7) SCI will consider whether periodicals are the new research results published in new periodicals, and if they can enrich the contents of the database they can be considered included (8) SCI will examine the degree of internationalization of periodical authors, editorial boards and editorial boards, while at the same time, SCI will strive for the balanced collection of different disciplines and regions (9) Based on citation database, SCI analyzes the importance and influence of different periodicals to realize the dynamic management of the database (10) For the new periodicals, analyze the journals of the main authors and members of the editorial board before publication in other journal papers to see whether there is a reference record. For the existing periodicals, the calculation of the impact factors. New journals will be selected every year to eliminate the use of less old journals	(1) Timeliness (2) International editorial practice (3) Full-text English (4) Peer-review process (5) Publication of moral statements (6) Publication format (7) Edit content (8) International diversity (9) Discipline and regional balance (10) Dynamic management	(1) Publication cycle (2) Editorial standardization (3) Academic integrity (4) Academic quality (5) International influence (6) Editorial quality (7) Capacity for sustainable development

C. English-language journals are prioritized, followed by journals published by European countries and the major journals of other countries.
D. Further key Ei requirements include the timeliness of journal publication and the degree of internationalization. Journals need to show academic contribution and original research.
E. Dynamic management: Ei follows a dynamic management model where journals are reviewed annually. Low-quality journals are removed and replaced by higher-quality journals.

Scopus

Scopus is a relatively new index, but it has a wide application including more than 2000 universities and research institutes around the world. Those students and researchers using Scopus are typically engaged in study and research work. Nearly all of the world's top laboratories, research institutes, and ranking agencies use Scopus for evaluation. Scopus is also used by innovation leaders in various industrial sectors for research and development. The database is accredited by the British government, and England's top four universities use Scopus as the only bibliometric tool under the 2014 Research Excellence Framework (REF) to assess the quality of research in higher education institutions. The Australian Scientific Research Council (ARC) uses Scopus to provide assessment support for national research and evaluation work.

Main Features. The Scopus database contains data from more than 100 countries, 5000 international publishers of more than 20 000 peer-reviewed journals, and covers all fields of science and technology, medicine, social sciences, and the arts and humanities. It contains 32 million abstracts from 1996 to the present, and 21.3 million abstracts from 1823 to 1996. Approximately two million records are added each year by means of daily additions. It represents the comprehensive integration of STM web page resources: 545 million science and technology web pages from the five largest patent organizations in the world and 25.2 million pages of content with a wide geographical distribution. More than half of the content comes from Europe, Latin America, and Asian countries.

Producer. Elsevier.

Target Journals. Scopus covers 313 subjects in 27 subject areas including medicine, physics, and mathematics.

Purpose of Evaluation. Scopus provides a fast and accurate retrieval of the full text and referenced information in real-time tracking. It offers the latest results from related research fields, hot topics, and institutions. Scopus aims to enhance the academic impact of individuals and institutions to promote global academic exchange. It also provides customized services for the data and analysis needs of government and evaluation agencies.

Evaluation Method. The Scopus Content Selection and Advisory Committee (CSAB) is an international and independent committee operated by a group of experienced international peer-review journal editors. It also includes publishing, bibliometric, and library science experts; these members were invited based on their desire to improve Scopus. At the same time, a key objective of CSAB is to ensure the recording of high-quality journals.

Selection Criteria. Stage One: Preselection criteria:

a. Timeliness
b. International editorial practices
c. The process of peer review
d. Author declarations
e. Reference specifications.

Stage Two:

1. Standard characteristics:
   a. Journal policy, journal operation
   b. The level of the peer-review experts
   c. The internationalization and diversity of editors and editorial boards, the internationalization and diversity of authors
   d. Geographical diversity of journals.
2. Content quality of journals:
   a. Academic influence
   b. The text is readable
   c. Publication history of at least two years.
3. Online availability of journals.

Evaluation Indicators. Scopus indicators include SJR, SNP, IPP, etc.

MEDLINE

MEDLINE is the bibliographic database of the National Library of Medicine (NLM) in the United States. It covers biomedical and life sciences topics critical to biomedical practitioners, educators, and researchers, and abstracts databases based on biomedical science. It includes the Medical Index (Index Medicus), Dental Literature Indexing (Indexes to Dental Literature) and the International Nursing Index. MEDLINE includes articles from 1966 to the present day, and the collected documents represent the highest quality in the global medical field. They also reflect the latest development and research results in biomedicine.

Main Features. The MEDLINE collection includes articles from more than 70 countries and regions, and includes more than 3400 different subject areas. Since 1946, close to 21.6 million records have been added, and it includes weekly updates and has a high proportion of English-language articles.

Producer. National Library of Medicine.

Target Journals. Core biomedical topics.

Purpose. To meet the needs of the world's researchers, healthcare workers, educators, managers, and students.

Evaluation Method. A technical review committee for the selection of documents was set up in 1988 to evaluate the content of journals, to co-opt new journals, and to include journals that do not conform to usual standards and requirements.

Selection Criteria.

A. Biomedical-oriented
B. Focuses on the quality of journal content, editing, and publishing
   - The content quality of journals: the academic level of journals is the primary factor of MEDLINE journals. The scientific aspects, timeliness, originality, and contribution of the article content are the important aspects.
   - Editorial quality: selection, method, and process of manuscript review, ethical and moral declarations, proof of conflict of interest, chart of production, corrections, and comments.
   - Publication quality: publication layout, printing, illustrations, and binding quality are also important aspects of the evaluation of journals.
C. Readers
   The focus is on the study of health professionals, including general researchers, caregivers, educators, managers, students, and other readers interested in healthcare.
D. Non-English journals
   Non-English journals and English-language journals have a consistent standard of assessment. Non-English journals should include English-language abstracts, and subject, author, unit, and chart following acceptable English-language standards.
E. Other
   Good quality journals will be included wherever they are published. Special attention is given to epidemics and endemic diseases. The diversity and internationalization of journal authors and the aca-

demic status of editorial commissioners will affect the collection of journals.

J-STAGE

J-STAGE (Japan Science and Technology Information Aggregator) was created by the Japan Science and Technology Agency in October 1999. The aim was to provide a network platform for academic journals. As the core agency responsible for Japan's Science and Technology Basic Plan, J-Stage undertakes the electronic and international internationalization of Japanese science and technology papers. J-STAGE focuses on Japanese science and technology papers, from their submission to their online publication. It aims to create a smooth process, from electronic submission to the publication processes and electronic special issues. J-STAGE not only contains submitted papers, but also other articles linked to the citation data, including video and audio, and appendices. The development of the J-STAGE database has networked the publications of Japanese academic and journal circles, and exemplifies the advantages of data integration. It helps to disseminate the achievements of Japanese science and technology and STM journals worldwide via this network. Thus, it aids in their effective use.

Main Features. J-STAGE was founded at the end of the last century, and focuses on scientific and technical journals, conferences, and reports. There is open access to all documents, with full access to the full text of science and technology journals. The J-STAGE is now in its 1808th issue, and 19% of the content is Japanese-language papers, 40% are in English, and 41% a mixture of the two.

Producer. Japan Science and Technology Agency.

Target Journals. The J-STAGE database includes the detailed categorization of journals. According to their respective disciplines, these are divided into 18 topics, covering, for instance, the natural sciences, humanities, social sciences, medicine, and engineering. Life science journals account for approximately 70% of the total.

Purpose of Evaluation. J-STAGE aids the electronic and networked processes of Japanese STM journals. The speed of the database is convenient for the publication of related papers, and the network of papers helps to accelerate the circulation of Japanese scientific and technological success. It contributes to the worldwide spread of Japanese scientific and technological achievements.

Evaluation Method. National institutions and other professional academic groups have developed the main approach.

Selection Criteria. Generally, important academic journals are recommended for entry into the database. Regular adjustments are made (these are generally small). The abstracts and titles of most of the papers included in the journal are in English. J-STAGE's selection criteria for journal content include:

- Reflect innovative viewpoints from particular fields
- Logically rational scientific papers that accord with scientific culture and ethics
- Real and reliable research resources and data
- High-quality writing skills.

20.3.2 Major National Evaluation System of Academic Journals in China

China's current evaluation system of journal retrieval includes the following databases: CSTPCD, Journal of Chinese Core Periodicals, Chinese Humanities and Social Sciences Core Periodicals, Chinese SSCI, and CSCI.

The CSTPCD indexes the *China's Science and Technology Core Journals*. CSTPCD includes Chinese publications (but excludes those from Hong Kong, Macao, and Taiwan) and currently references 2312 Chinese and 71 English-language journals. It contains 2383 core Chinese STM journals (i. e., Chinese STM source journals), which are classified into two or three subjects according to 153 subjects in 10 fields (e. g., natural science synthesis and neo-Confucianism).

A total of six editions of CSTPCD were published between 1992 and 2014. Selected journals were assessed and qualitatively evaluated, and nearly 2000 core journals were selected from journals published in China. The database contains seven sections (excluding interdisciplinary repetition) and the 2014 edition contains 1983 journals. The database has a great influence on the society via its provision of a reference for the evaluation and ordering of journals for information departments and the management policy of administration departments.

The 2014 edition contains 733 journals concerned with humanities and social sciences. The journals are distributed in 23 categories and are divided into authoritative journals, core journals, and extended journals in different subjects.

The CSSCI was first established in 1998 and four editions have been published. The latest version (2014–2015 edition) concerns, for example, management, Marxism, and another 25 disciplines. It contains 533 journals, including the largest number of comprehensive journals. Furthermore, it includes comprehensive social sciences and 120 efficient comprehensive journals, accounting for 22.5% of the total number of journals.

Table 20.3 CSTPCD features and structure

Raw materials	Deconstruction	Reconstruction
(1) CSTPCD mainly includes national, provincial or regional outstanding periodicals in the field of natural science (2) The classification of 153 subjects and 10 subjects including natural science synthesis and CSTPCD (3) The citation report of Chinese STM periodicals started in 1988, with the latest edition in 2014, and a yearly report (4) CSTPCD includes Chinese publications (excluding Hong Kong, Macao): 2312 Chinese periodicals and 71 English periodicals, a total of 2383 ...	(1) The field of natural science (2) STM periodicals (3) Academic achievements (4) Involving national science (5) Interdisciplinary (6) Update frequency (7) Time range ...	(1) Major disciplines and interdisciplinary classifications (2) Domestic influence (3) Journal attributes (4) Update efficiency and timeliness

The CSCI 2015–2016 includes 1200 source journals, referencing Chinese (1006) and English-language (194) journals.

China Scientific and Technical Papers and Citations Database (CSTPCD)

Main Features. CSTPCD includes science and technology journals in the field of natural science, with a large number of journals and covering a wide time span. It includes interdisciplinary journals for multidisciplinary evaluation.

Producer. ISTIC was commissioned in 1987 by the Ministry of Science and Technology to undertake statistical work.

Target Journals. State-level academic journals, CAS journals, academic journals from key universities, and all natural and social sciences journals.

Purpose of Evaluation. CSTPCD was first published in 1998, with two versions: the core edition and the expanded edition. The latest edition was published in 2014. It is a journal evaluation tool for a vast number of scientific and technical personnel, journal editors, and scientific research managers, enabling them to quickly, accurately, and scientifically select and use journals.

Evaluation Method. Mainly uses a multiple index evaluation system, a combination of quantitative and qualitative methods, and quantitative methods (using a series of bibliometric indexes) to appraise journals.

Selection Criteria. The journals covered by CSTPCD should be peer-reviewed academic journals that have been published for more than 2 years, and whose indicators ranked in the forefront of the discipline. They should be in line with academic publishing norms and meet the publishing integrity and ethical requirements.

Evaluation Indicators. The CSTPCD uses two evaluation methods: single-index evaluation and comprehensive-index evaluation. Specific indexes include citation frequency, impact factor, important database collection, and comprehensive evaluation score.

According to the upper introduction and study materials about CSTPCD, we separat those raw materials into deconstruction items and re-organize them as reconstruction item as similar structure in order to comparative study different evaluation system (Tables 20.3 and 20.4).

A Guide to the Core Journals of China

Main Features. *A Guide to the Core Journals of China* (2014) contains 1982 core journals, published every four years from 1992 to 2011, and then again in 2014.

Producer. Developed by Peking University Library.

Target Journals. Covers a wide range of disciplines (Table 20.5).

Purpose of Evaluation. To provide reference for the library's journal purchases and Chinese journal collection, and to facilitate readers' access and authors' contributions.

Evaluation Method. Quantitative and qualitative comprehensive analysis and evaluation, based on bibliometrics, the use of evaluation indicators for the domestic publication of Chinese journals for statistical analysis, and the use of expert opinion.

Selection Criteria.

A. Indicator system: according to the law of Brinell, bibliometric statistics are used to screen the list of core journals from different disciplines.
B. Division of disciplines: overview using the medium-map method to divide the subject, based on subject size, number of journals, journal quality, and other factors that constantly modify the discipline.
C. Core journals are selected based on three aspects: the journal is considered a core journal, it is representative, and it is practical.

Table 20.4 CSTPCD description and structure

Raw materials	Deconstruction	Reconstruction
(1) CSTPCD includes mathematics, physics, medicine and other fields (2) CSTPCD includes STM and technical periodicals reflecting the development of scientific and industrial technologies (3) CSTPCD requires the periodical to conform to the description specification, such as the unified CN number, complete bibliographic information, etc. (4) The publication of CSTPCD from the peer-review process (5) Strict publication cycle of CSTPCD Source Journal (6) CSTPCD will consider whether the periodicals are indexed in SCI, SCIE, Ei, CA, and so on, as well as famous large-scale search systems (7) The social influence and academic status of the source periodicals. Academic reputation can attract high-quality papers at home and abroad. Excellent academic research journals at the local level (8) CSTPCD-included periodicals need to follow the international and domestic editing practices, and need to meet the normative requirements of the journal: complete bibliographic information and so on (9) Academic content should reflect the latest achievements in the field of science and research projects of major research funds. Should have the national authoritative expert composition of the editorial committee. Rigorous academic accreditation of periodicals (10) The evaluation principle must be a combination of qualitative and quantitative (11) To ensure the balanced collection of periodicals in different disciplines, especially new disciplines or high technology disciplines, and to take care of the new development area periodicals to ensure the integrity of the area	(1) Subject area (2) Domestic and international editorial practice (3) Editorial authority (4) Timeliness of publication (5) Domestic and foreign retrieval system included (6) Academic quality (7) Comprehensive assessment (8) Special consideration (9) Discipline and regional balance (10) Dynamic management	(1) Subject-biased (2) Editorial standardization (3) Editorial quality (4) Academic quality (5) International influence (6) Capacity for sustainable development

Table 20.5 The distribution of journals in *A Guide to the Core Journals of China*

Volume	Field	Number of periodicals
1	Philosophy, sociology, politics, law	274
2	Economics	155
3	Culture, education, history	311
4	Natural science	344
5	Medicine, health	250
6	Agricultural science	135
7	Industrial technology	514
	Total	1983

Evaluation Indicators. The seventh edition (2014) is based on 12 evaluation indicators such as the amount of article Full-Text requested in library system, the amount of article covered by important databases, the citation index of paper, the index of mutual citation etc.

Chinese Humanities and Social Sciences Core Journals Database (CHSSCD)

Main Features. The CHSSCD includes more than 700 journals on humanities and social sciences, and was established in 1996.

Producer. Chinese Academy of Social Sciences (CASS) Literature Information Center.

Target Journals. Humanities and social sciences journals (Table 20.6).

Purpose of Evaluation. To optimize the use of journals and literature resources for scientific research, to provide reference for journal evaluation, scientific research performance evaluation, scientific research management, and talent selection in scientific research.

Evaluation Method. Combination of quantitative evaluation and expert qualitative evaluation.

Selection Criteria. The 2014 report on the indicator system of journal evaluation creates three categories: attractions, management power, and influence. These are described below:

- *Attractions*: Academic reputation (award status, peer review), the inclusion of other domestic databases, the diversity of authors, and the quality of papers
- *Management power*: Journal orientation, academic ethics, editorial staff quality, editorial standardization, publishing norms, and networking.
- *Influence*: Academic quality of journals, internationalization of editorial board, social and international influence of journals, etc.

Table 20.6 Grade distribution for China's humanities and social sciences journals 2014

Rank	Field	Number of periodicals				
		Total	Top	Important	Core	Extend
1	Law	32	1	2	19	10
2	Management	25	1	2	12	10
3	Environmental science	5	0	1	3	1
4	Pedagogy	33	1	2	20	10
5	Economics	109	1	3	63	42
6	Archeology	18	1	2	9	6
7	History	32	1	2	21	8
8	Marxism	14	1	2	8	3
9	Ethnology and cultural studies	27	1	2	15	9
10	Human geography	12	0	1	7	4
11	Sociology	15	1	2	6	6
12	Physical education	14	0	1	7	6
13	Statistics	4	0	1	2	1
14	Library, information and archival science	31	1	2	16	12
15	Literature	24	1	2	14	7
16	Psychology	7	0	1	4	2
17	Journalism and communication science	11	1	1	5	4
18	Art	16	1	1	9	5
19	Linguistics	32	1	2	19	10
20	Philosophy	15	1	2	7	5
21	Politics	68	1	2	42	23
22	Religion	2	0	0	2	0
23	Comprehensive humanities and social sciences	187	1	4	120	62
Total		733	17	40	430	246

Evaluation Indicators. The 2014 edition uses citation frequency, the difference between the two-year and five-year impact factors, the article funding ratio, article downloads, and so on. This database is able to determine the current situation of journals in China.

Chinese Social Sciences Citation Index (CSSCI)

Main Features. The CSSCI (2014–2015) contains 533 humanities and social sciences journals; these have been selected four times since 1998.

Producer. Nanjing University.

Target Journals. Humanities and social sciences journals, divided into 25 fields (Table 20.7).

Purpose of Evaluation. It provides reference and help for academic evaluation, performance, management, and research on humanities and social sciences.

Evaluation Method. A combination of quantitative and qualitative methods.

Selection Criteria.

A. Principles of inclusion: the principle of prioritizing quality journals, the control of the total number of journals, and the use of quantitative and qualitative evaluation for regional and discipline balance, and dynamic high and low management.
B. Editorial specifications: full information provided with the academic norms of reference and literature notes.
C. Time and timeliness of the publication: Publishing frequency, deadlines met, published continuously for five years, the extension of more than two months of published journals belonging to the editors (journals that do not follow publishing norms are not selected).
D. Disciplines and sources of publications: a main focus on humanities and social sciences academic papers, academic reviews, and other original academic literature. The number of source journals is limited to 20% of the total academic journals of humanities and social sciences in China.
E. Citation factors, total citation frequency and other indicators, with different weights given to each for the quantitative evaluation of journals.

Evaluation Indicators. A Chinese Social Sciences Research Evaluation Center, Nanjing University, (2016–2017) publication outlines the quantitative evaluation of journals by citation quantity and other factors.

Table 20.7 The subject distribution of CSSCI

Field	Number of periodicals
Management	29
Marxism	16
Philosophy	12
Religion	3
Linguistics	23
Foreign literature	6
Chinese literature	16
Art	21
History	26
Archeology	7
Economics	73
Politics	32
Law	21
Sociology	10
Ethnology and cultural studies	14
Journalism and communication science	15
Library, information and philology	20
Pedagogy	36
Physical education	10
Statistics	4
Psychology	7
Comprehensive social sciences	50
Humanities, economic geography	7
Environmental science	5
Comprehensive Journal of Colleges and Universities	70
Total number of periodicals	533

China Science Citation Database (CSCD)

Main Features. The CSCD includes journals in the following areas: natural science, engineering technology, medicine, and other fields of science and technology. It includes thousands of journals, ranging from 1998 to the present day, and includes 300 000 articles, and nearly 17 million citations. Source journals are selected every two years.

Producer. Chinese Academy of Sciences Document Information Center (now known as the CAS National Library).

Target Journals. The CSCD focuses on basic scientific research in the field of natural science.

Purpose of Evaluation. It provides the basis for selecting source journals and evaluating STM journals for Chinese scientific citation databases.

Evaluation Method. Comprehensive evaluation method combining quantitative statistics and expert evaluation.

Selection Criteria.

A. Theoretical basis: based on literature concentration and the discrete law of Brinell.
B. Editing and publishing norms: has both an ISSN and CN, two standard publication numbers, and the journals must conform to standard journal descriptions.
C. Scope of disciplines: covering the fields of mathematics and physics, while paying attention to the collection of basic research, academic, theoretical, and leading-edge journals.

Evaluation Indicators. The index uses more than ten factors including impact factor, quality index, thesis utilization index, and mutual index.

20.3.3 Comparison of International and Chinese Journal Evaluation Systems

By summarizing the data of different journal evaluation systems, we can outline and refine the different attributes of various domestic journal evaluation systems (Table 20.8).

Main Features

The characteristics of the systems and the target journals can be divided into five categories: subject, country, the number of journals, the time span, and updating efficiency (Table 20.9). From the point of view of the distribution of disciplines, foreign journal evaluation systems contain major areas of difference. For example, SCI, the major areas of Ei, Scopus, MEDLINE, and J-STAGE concern natural sciences, social sciences, engineering technology, and biomedicine. Furthermore, foreign journal evaluation systems focus on those jour-

Table 20.8 Key characteristics of the main domestic and international journal evaluation databases

	Owner	Headquarters
SCI	Thomson Reuters	United States
Ei	Elsevier	United States
Scopus	Elsevier	Netherlands
MEDLINE	National Medical Library of America	United States
J-STAGE	Japan Science and Technology Revitalization Agency	Japan
CSTPCD	China Institute of Science and Technology Information	China
GSJC	The Library of Peking University	China
CHSSCD	Literature Information Center of Cass	China
CSSCI	South Social Science Research Center	China
CSCD	National Library of CAS	China

Table 20.9 Key features of international and national journal evaluation databases

	External features	Key disciplines	Number of periodicals	Operating since	Update time
Main international evaluation system	SCI	Natural science	>12 000	1900	Yearly
	Ei	Engineering technology	3639	1884	Weekly
	Scopus	Social	>20 000	1823	Daily
	MEDLINE	Medical	>3400	1946	Weekly
	J-STAGE	Natural science	>1000	1999	
Main evaluation system in China	CSTPCD	Natural science	2383	1987	Yearly
	GSJC	Wide range of disciplines	1982	1992	Three years
	CHSSCD	Human society	>700	1996	Yearly
	CSSCI	Human society	533	1998	Biennially
	CSCD	Natural science	>1200	1998	Biennially

nals that embody the characteristics of the subject and contribute to journal focus. For example, Ei focuses solely on engineering, and includes all journals with an engineering technology scope. From an evaluation perspective, the foreign evaluation system is discipline-oriented, which is helpful to find suitable data for analysis and research in different disciplines. Therefore, the distribution of journals reflects that the foreign journal evaluation system and databases pay attention to the subject matter and the specialty of the journals. The field of focus of the journal is an important aspect of foreign evaluation databases. China's journal evaluation systems include major areas of difference: CSTPCD and CSCI focus on natural science, CHSSCD (CASS) and CSSCI include humanities and social sciences, and GSJC (Peking University) journals are covered extensively. From the perspective of journal discipline, the foreign evaluation system places a greater emphasis on the subject characteristics of journals, and it focuses on the journals with academic contributions and clear characteristics.

Regarding the scope of coverage, SCI, Ei, MEDLINE, and Scopus cover more than 70 countries and regions, and significantly more than than J-STAGE's 3400 journals (mainly from Japan). Furthermore, SCI, Ei, and MEDLINE focus on Western countries. Scopus' content is more balanced in terms of geography, with more than half of the content coming from Europe, Latin America, and Asia. Thus, Scopus attracts more countries and regions to use and search the index. The journal evaluation system for different journals will consider the balance of the region and the subject. China's journal evaluation system only contains important journals from China. Because its geographical scope is narrow, the number of journals within China's journal evaluation systems is obviously less than that of the foreign journal evaluation systems. Therefore, it is worth considering whether it is necessary to enlarge the scope of China's journal evaluation system and to enhance its international influence.

Regarding the history of the various systems, international evaluation systems such as SCI and Ei have been operating for some time now. Ei has a long history of 130 years, while SCI was established 60 years ago. In contrast, China's index database and evaluation system were established just 30 years ago. The period of operation reflects the maturity degree of the journal evaluation system: the longer the history of the system, the greater the verification of the system's practices, and the greater the opportunities to improve and fine-tune the process. This can also occur via international influence and use. Therefore, China's journal evaluation system may be able to draw lessons from the development process of overseas journal evaluation systems, and provide a reference for the healthy development of its journal evaluation systems.

Regarding the frequency of updating, foreign journal evaluation systems are updated frequently. For example, SCI, Ei, and MEDLINE data are updated weekly, and Scopus data are updated daily. These updates are quick, ensuring the data remain relevant. Regarding the publication of citation reports or core journals in the journal evaluation system at home and abroad, JCR and CHCR publish once a year and *A Guide to the Core Journals of China* is now published once every three years. The CASS humanities and social sciences database is published annually. Thus, regular updates occur in both the domestic and foreign journal evaluation systems, as well as the dynamic management of the data and core journals, which reflects the continuous increase in user demand. The evaluation system continues to update the data, and reflects the sustainable development ability of the evaluation system. Therefore, China's journal evaluation system can also be combined with its own resources to provide users with more comprehensive search and data services to promote the sustainable development of journal evaluation.

Subject Characteristics

The main evaluation systems include ten domestic and international evaluation systems from three companies, three libraries, and four research institutions. Sixty percent of the foreign evaluation systems are owned by companies. In contrast, China's main journal evaluation systems include three scientific research institutions, and two universities or research institutions. Thus, China's journal evaluation systems are concentrated in research institutions and libraries.

As stated above, international journal evaluation systems are mainly owned by companies, while China's are mainly owned by scientific research institutions. Corporate enterprises are profit-oriented, so pay greater attention to user needs. Scientific research institutions generally have a clear direction and aim (i. e., academic research), and pay more attention to academic results and so on. Both ownership styles have their advantages and disadvantages.

Objective and Method of Evaluation

The index results are largely read by scientific research personnel, management departments, and librarians. For scientific researchers, the results of the journal evaluation provide a standard for evaluating the performance of scientific research. Furthermore, managers will use the evaluation results for performance appraisals and the development of management journals. Librarians will use the results of journal evaluation to determine which journals they will stock in their collections. Therefore, from the reader level, the use of the journal evaluation results regarding the extent and degree of accreditation can also reflect the results of the journal evaluation in terms of rationality, scientific merit, and influence.

The evaluation method of the journal appraisal system is based on quantitative evaluation and expert qualitative analysis, which is based on bibliometrics. That is, objectivity, the measurement and comparability of scientific research results from different perspectives are useful in the objective, fair, and reasonable evaluation of academic journals.

Selection Method

Based on the descriptions and standard main evaluation system at home and abroad, we continue to process the content of the various systems. Table 20.10 shows our results. Overall, the foreign evaluation systems pay more attention to, for example, journal content, edition, publication quality, and degree of internationalization. This can be seen in the fact that the foreign journal evaluation systems largely request the paper to have an English title, abstract, and so on. Similarly, other import factors include the journal's internationalization and diversification requirements, the nationality of the author, and the degree of internationalization of the editorial board and international evaluation experts. In China, the main evaluation systems pay attention to the categorization of the journal's subject range, the number of journals included in key domestic and foreign retrieval systems, editorial authority, and the influence of the journals. A further important point includes whether the journals conform to national standards (e. g., has a CN).

Based on the criteria and principles of the evaluation systems, we will identify the criteria and principles of the different evaluation systems, and summarize the evaluation systems of various journals in combination with their key characteristics (Table 20.11).

On the whole, foreign journal evaluation systems pay greater attention to the following aspects: journal influence, academic quality of journals, standardization of journal editors, timeliness of journal publication, author declarations, originality, dynamic management, editorial and author diversity and nationality, English-language content, discipline and regional balance, and peer reviews. In contrast, Chinese journal evaluation systems focus on: journal influence, academic quality of journals, the standardization of journal editors, timeliness of journal publication, and dynamic management. It can be seen that the domestic and foreign journal evaluation systems attach great importance to journal

Table 20.10 Purpose of evaluation of domestic and foreign journal evaluation databases

	Purpose of evaluation (object-oriented)	Evaluation method
SCI	Research personnel, management, etc.	Qualitative and quantitative combination
Ei	Scientific researchers, engineers and technicians	Qualitative and quantitative combination
Scopus	Academics, researchers, government administrations, etc.	Content Selection & Advisory Committee (CSAB)
MEDLINE	Nursing staff, healthcare providers, educators, researchers, academics and administration	Technical Review Committee on Document Selection (LSTRC)
J-STAGE	Researchers, academics, etc.	National Society and other professional academic groups recommend
CSTPCD	Technicians, editors and managers	Qualitative and quantitative combination
GSJC	Librarians, readers, research workers	Qualitative and quantitative combination
CHSSCD	Journal reviewers, researchers and management	Qualitative and quantitative combination
CSSCI	Provide reference for evaluation, performance appraisal, management and research	Qualitative and quantitative combination
CSCD	Provide basis for periodical evaluation	Qualitative and quantitative combination

Table 20.11 Selection criteria standards and principles included in domestic and international journal evaluation systems

	Standards and principles
SCI	Timeliness, editorial specification (English abstract, etc.), full text English, peer review, publishing moral statement, author and editorial board international diversity, discipline and regional balance, dynamic management, etc.
Ei	Editorial standards (English abstracts, etc.), priority of English-speaking countries, timeliness, internationalization, originality, academic contributions, dynamic management, etc.
Scopus	Timeliness, editorial standardization (English abstracts, etc.), peer review, moral statement, journal policy, academic level of the accreditation committee, international diversity of editorial board and authors, regional and discipline balance, journal influence and quality, etc.
MEDLINE	Academic level, timeliness, originality and contribution degree of the periodical, moral statement, editorial norm, editorial board and author's international diversity, regional and discipline balance
J-STAGE	Take recommendations as the main basis
CSTPCD	Editorial norms, authoritative editorial board, timeliness, the collection of important retrieval systems at home and abroad, the social and academic influence of periodicals, the balance of regions and disciplines, etc.
GSJC	Based on the quantitative selection of the core list, the use of middle-map method to divide the size of the subject; periodicals are representative and practical
CHSSCD	Academic quality of periodicals, collection of other domestic databases, diversity of authors, academic ethics, editorial norms, academic level of editorial board and international diversity, periodical influence, etc.
CSSCI	Journal quality, regional and discipline balance, editorial norms, timeliness, original innovative academic literature, etc.
CSCD	Editorial standardization, focus on basic research, academic and theoretical frontier periodicals, etc.

influence, academic quality, the standardization of journal editors, and the timeliness of publication. It can also be seen that dynamic management is key. It is only via dynamic management that an evaluation system will have the vigor to constantly develop, and to meet the increasing user demand.

There are also some differences in journal evaluation systems. For example, compared with China's journal evaluation systems, foreign systems pay more attention to author declarations and originality, the diversity and internationalization of the editorial board and author, English-language content, and peer-review processes. Regarding author declarations and originality, the governing principles of the journal evaluation systems in foreign countries concern the issue of academic ethics and ethical publishing. Thus, articles will be removed from the relevant database should they breach any ethical standards. That is, users can see an article has been revoked. Regarding ethics and conflicts of interest, in 2013 MEDLINE introduced stricter requirements for journals (in terms of application and after acceptance). Regarding the internationalization and diversification of the editorial board and authors, the foreign evaluation systems basically stipulate the international diversity of editorial board and the authors. The diversity of the authors is beneficial to the global influence of journals, and the international diversity of the editorial board is more advantageous when checking the quality of journal papers and improving academic quality. Concerning journal abstracts, it is generally required that the journals need to have English-language abstracts, which is not only beneficial to the unification of data form, but also to the use and dissemination of knowledge. With regard to expert review, foreign evaluation systems generally consider peer review to be an important guarantee to ensure journal quality. In addition, foreign systems have a high degree of network and computerization, and for MEDLINE, journal publication quality is an important selection factor. Generally speaking, foreign journal evaluation systems have a high quality guarantee in their journal selection, and fully embody internationalization. These systems pay close attention to information dissemination, and seek consistency.

Consistent with the key features outlined above, Ei and MEDLINE require that journals have the desired subject characteristics and the necessary academic requirements. At the same time, it can be seen from the standard of journals, both the subject and journal subject are prerequisites for Ei, and all the journals are engineering technology journals. Furthermore, the features of Ei and the selected journals are notable. For interdisciplinary journals, different journal evaluation systems have different ways of dealing with them, including SCI and CSTPCD (2015 edition). These databases calculate the journal index under different disciplines and highlight the subject characteristics of the journal.

In terms of selection criteria, the standards and principles can be divided into four categories: journals, editors, authors, and judges (Table 20.12). China's journal evaluation systems and those abroad attach great importance to the journal's core value, including the journal's influence, academic quality, editorial quality, and ethics. As journal editors and managers, the editorial board of foreign systems evaluates the internationalization and academic level of the editorial board.

Table 20.12 Summary of standard principles of journal evaluation systems

Standards and principles	SCI	Ei	Scopus	MEDLINE	J-STAGE	CSTPCD	GSJC	CHSSCD	CSSCI	CSCD
Journal influence	✓	✓	✓	✓		✓	✓	✓	✓	✓
Academic quality	✓	✓	✓	✓		✓	✓	✓	✓	✓
Editing specifications	✓	✓	✓	✓		✓	✓	✓	✓	✓
Timeliness	✓	✓	✓	✓		✓	✓	✓	✓	✓
Moral statement and originality	✓	✓	✓	✓				✓	✓	
Dynamic management	✓	✓	✓	✓		✓	✓	✓	✓	✓
Editorial international diversity	✓		✓	✓		✓				
Diversity of authors	✓	✓	✓	✓						
English (title, abstract, etc.)	✓	✓	✓	✓				✓		
Regional discipline balance	✓		✓	✓		✓			✓	
Other important retrieval systems		✓	✓			✓		✓		
Peer review	✓	✓	✓		✓	✓				
Degree of contribution to the discipline		✓		✓						✓
ISSN or CN	✓	✓	✓	✓		✓	✓	✓	✓	✓

Additionally, the editorial board is an important factor in a journal's core value. The author is the main contributor to the content of the journal, but unlike the editorial board, the author is not fixed. To a certain extent, the author is also the factor that affects the journal's value, so foreign journal evaluation systems will also consider the diversity of the journal authors. The evaluation is more like foreign aid to the journal—it provides professional guidance and recommendations for the editorial board. Furthermore, it is advantageous to safeguard the journal quality. It can be seen both in the database descriptions in China and abroad that the journal, editorial board, and other actors are investigated and evaluated. Thus, it is worth considering whether these tasks could be performed by other *evaluators*.

Summary

From the above-mentioned domestic and foreign journal evaluation systems, we can see that any journal evaluation and screening system is based on bibliometrics. Through the deep discussion of Bradford's laws, we can see that the selection and evaluation of source journals is carried out under a framework of their respective evaluation purposes, evaluation methods, key features, and evaluation criteria, each with applicable scope and evaluation characteristics.

Foreign evaluation systems (e. g., SCI, Ei, and Scopus) are located all over the world, and include the key journals from many countries. Thus, the journals included in the evaluation systems of foreign major journals reflect the international influence of the journals to some extent. For Chinese journals, many of these journals with Chinese characteristics (e. g., differences in language and editing norms) have not been included in the key evaluation systems. To some extent, this can be explained by the fact that ranking of China's STM journals in the international arena still needs to be improved. Therefore, certain aspects of the foreign evaluation systems are not completely suitable for China. At the same time, the foreign systems offer considerable experience from which China can learn from in developing its own evaluation system.

Based on the induction and refining process of the domestic and foreign journal evaluation systems, we find that the journal evaluation systems are largely concerned with the following issues:

1. The key characteristics of journals (professional characteristics or academic contributions)
2. Influence of journals
3. Dynamic management of journals (sustainable development)
4. Academic honesty and morality in publishing
5. Diversity and internationalization of editorial board
6. Diversity of authors
7. The balance of areas and disciplines
8. Standardization of editors
9. Readability (e. g., English-language abstract).

20.4 Comparative Study of International and National Evaluation Indicators of Academic Journals in China

20.4.1 Popular Evaluation Indicators of International Academic Journals: JCR

Comparatively speaking, foreign journal evaluation systems adopt a periodical journal evaluation index that is based on Garfield's citation analysis to statistically appraise citation data. Using bibliometrics and a combination of qualitative and quantitative methods to evaluate and select journals, JCR evaluations mainly include the annual number of published documents, citation frequency, and the calculation of impact factors. JCR has recently introduced evaluation indexes for feature factors and influence scores. Generally speaking, JCR is considered an authority on citation data.

The JCR index, providing a statistical and comparative analysis of Chinese journals and other major journals, includes country of origin data (Tables 20.13–20.15). The total number of journals included in the JCR (Science Edition) increased considerably between 2012 and 2014; so too has the number of Chinese journals. In contrast, there has been a decline in the number of journals from France, Russia, Japan, and India. The number of journals from the United States, Britain, and Germany is significantly higher than those from other countries. The number of Q1, Q2, and Q3 journals from China increased in the 2012–2014 period. Furthermore, China enjoys the highest growth rate for journal number, followed by South Korea, while France has largest decline.

A journal can focus on many subjects, and the impact factor may be different for different disciplines. If some journals have multiple subject areas, then the impact factor is calculated for each subject area. Thus, makes the number of journals in four sections of the table more likely than the total number of journals. The term *China (Exclude Duplicates)* means that if a journal focuses on multiple disciplines, only the best area is considered, at which time, the sum of the four areas equals the total number of journals.

Regarding the distribution of impact factors, the distribution of the number of journals in the four regions is more balanced, and the division of German journals represents the global situation. The distribution of U.S. and English journals differs from that in other countries, with the number of journals in the two countries decreasing with the increase in the number of partitions (e. g., Q1 and Q2 journals, Q3 journals, and Q4 journals). The opposite is true for the distribution of journals in other countries, including China. The proportion of Q1 journals in Britain, the United States, and Germany is much higher than that of other countries, and China's Q1 journals are second only to those three countries.

Regarding the number of English-language journals, the number of English-language journals from India remained stable, and the number of English-language journals from Japan decreased. The number of English-language journals from other countries increased. Furthermore, the number of English-language journals in the Q1, Q2 and Q3 divisions increased between 2012 and 2014. The three countries with the increase highest rate of English-language journals are South Korea, Brazil, and China; Japan shows the fastest decline.

Most of the journals entering the Q1 division in China are English-language journals. A total of 18, 19, and 21 Chinese STM journals were included in the JCR in the above three-year period, respectively, and the number of journals has increased gradually in recent years. Our analysis of these journals also shows that a number of Chinese STM journals have made it into the Q3 and Q4 divisions; 94.4% in 2012, 94.7% in 2013

Table 20.13 Number of journals included in JCR 2012

Geographic region	Number of periodicals	Q1	Q2	Q3	Q4
Global	8471	2287	2646	2616	2582
United States	2825	1046	1002	888	546
United Kingdom	1710	631	655	523	282
France	195	20	36	52	109
Germany	564	139	179	163	200
Russia	150	3	9	22	123
Japan	239	18	54	100	105
Korea	90	7	15	37	42
Brazil	102	0	8	25	73
India	105	1	7	22	82
China	156	11	33	60	71
China (excluding duplicates)	156	11	31	53	61

Table 20.14 Number of journals included in JCR 2013

Geographic region	Number of periodicals	Q1	Q2	Q3	Q4
Global	8539	2309	2649	2661	2592
United States	2875	1021	982	906	610
United Kingdom	1747	672	654	520	288
France	186	20	39	43	101
Germany	563	143	183	171	181
Russia	149	2	6	27	122
Japan	236	18	48	96	108
Korea	91	1	16	46	40
Brazil	108	0	11	29	73
India	98	0	8	23	74
China	167	14	42	59	78
China (excluding duplicates)	167	14	36	51	66

Table 20.15 Number of journals included in JCR 2014

Geographic region	Number of periodicals	Q1	Q2	Q3	Q4
Global	8618	2354	2670	2679	2638
United States	2894	1040	982	900	619
United Kingdom	1784	672	649	514	323
France	174	18	46	43	89
Germany	578	150	195	167	194
Russia	148	3	5	32	117
Japan	234	15	50	94	114
Korea	102	4	20	51	46
Brazil	106	0	4	34	73
India	97	1	5	17	77
China	179	23	44	69	71
China (excluding duplicates)	179	23	38	57	61

and 90.5% in 2014. However, these rates show that the citation rates of Chinese STM journals are not considered influential. The reasons for this include the language of the journals, the fact that the main readers and citations of Chinese journals are concentrated in China (which results in a domestic rather than an international influence), and that it is difficult to get the attention of (and hence citations from) international academia.

20.4.2 National Journal Evaluation Systems in China

The comparison results of the evaluation methods of various systems show that the current journal evaluation systems generally use a comprehensive evaluation method that combines both qualitative and quantitative approaches. Regarding quantitative evaluation, each evaluation system has a set of indicators. This chapter focuses on the different indicators, highlighting similarities and differences. Looking at the different indicators, we will identify design principles, algorithm design, and the applicability and limitations of the indicators via comparative analysis. We have already identified the key characteristics, standard profiles, main principles, and evaluation indexes for the main evaluation systems in China and elsewhere. Thus, we have analyzed these systems based on various system-level indexes. We continue the analysis of these evaluation systems in Table 20.16.

The evaluation indicators used for the academic journals include impact factors and the total frequency of citation. These indicators directly reflect the academic quality of the journal papers cited and the level of influence. The article funding ratio, web downloads, and annual index are further evaluation indicators in many evaluation systems. At present, the journal evaluation systems basically select journals by looking at their overall reputation and long-term practices, as well as their level of authority and representation. Both qualitative and quantitative methods are used to conduct journal evaluations. Influenced by foreign journal evaluation systems, Chinese journal evaluation indexes have been revised to some extent from traditional evaluation indexes. That is, the impact factor index has produced further evaluation indexes for journal evaluation.

Table 20.16 Main evaluation indexes of Chinese journal evaluation systems

Evaluation system CSTPCD Evaluation indicators	GSJC	CHSSCD	CSSCI	CSCD
Overseas papers rate	Web downloads	Two-year and five-year impact factors	Citation amount	Article influence score
Number of institutions distributed	The amount of article Full-Text requested in library system	Cited frequency	Impact factor	Eigenfactor score
Funded papers rate	Cited quantity	Funded papers rate	Total cited frequency	Index of mutual indexing
Year indicator	Total quoted times	Year indicator		Diffusion factor
Source literature quantity	Quoted paper rate	Paper download		Paper use index
Average number of citations	Indexed by important retrieval systems	Subject expansion indicators		Other cited frequency
Average number of authors	Index of mutual indexing	Subject impact indicators		Other influencing factors
Other citation	Funded papers rate	Important databases		Excellent index
Rate of document selection	Paper cited index			
Quote half-life	Other cited			
Impact factor	Other influencing factors			
Important databases	Impact factor			
Total cited frequency				
Total score of comprehensive and evaluation				

The impact factor index includes a two- and five-year impact factor. The impact factor and the impact of its peak period are based on the impact factor in the computation process difference and different calculations.

The total score of the comprehensive evaluation of Chinese STM journals (hereinafter referred to as the *comprehensive evaluation score*) is based on the comprehensive evaluation indicator system of Chinese STM journals. The relative position score of each journal in the discipline is calculated according to the extreme value of each journal index in each subject category. The annual CHCR (core edition) of the *China Science and Technology Journal* published by CSTPCD regularly publishes the scientific measurement indexes of the statistical source journals of Chinese STM papers [20.6, 7]. Based on this index, a comprehensive evaluation indicator system for Chinese STM journals was implemented in 1996 and further developed. Each index in the evaluation indicator system is not given equal weight when calculating the total score; that is, different indexes are given different weights. The initial index weight distribution was determined by Delphi expert investigation and AHP. Subsequently, based on further research and feedback, new measurement indexes were continuously added. To ensure that feedback on the evaluation results can be properly given and received, ISTIC has held more than 20 expert seminars for scientific metrology experts, scientists, science and technology and journal management experts, and scholars to participate in evaluation, establish indicators, and to determine the weight of updates and adjustments. Thus, they contribute to the creation of the current set of scientific and technological journals to improve evaluation systems and methods.

The total score of comprehensive evaluation is based on the principle of scientific metrology, carefully considering the indexes of the evaluated journals. On the basis of widely solicited opinions, a comprehensive evaluation indicator system for Chinese STM journals is used to classify, divide, and assign different weights to journal indexes. The weighted scores of each index are collected to obtain the total score of the evaluated journals and the ranking within the scope of the subject and in the range of journals: the higher the value of the comprehensive evaluation index, the higher the comprehensive academic quality and impact level of the journal in its subject area.

Because the evaluation aim, scope, and perspective differ among journal evaluation systems, the index range and corresponding weight distribution are frequently adjusted. Taking the 2015 CHCR (core edition) as an example, the total score of the comprehensive evaluation, looking at the core of the main citation frequency (TC), the core impact factor (IF), the core six indexes of this index (OT), the article funding ratio (NT), citation rate (RE), and open factor (OP), is based on the relative position of the evaluated journal in its subject. These indexes are integrated according to certain weight coefficients.

20.4.3 Indicator Design Method for Academic Journals

Based on the statistics of the evaluation index distribution of the main evaluation systems at home and abroad, the indexes of three or more evaluation systems are summarized. Among them, indexes can be merged; for example, the citation index and mutual index are merged to create citation frequency, and the two-year and five-year impact factor become the impact factor index. The most commonly used indicators in the different evaluation systems are as follows: citation frequency, impact factor, the annual index, the thesis funding ratio, download volume, and the key retrieval system. Among them, the traditional evaluation indicators include citation frequency and impact factor. Citation frequency and impact factor are widely used, appearing in all evaluation indicator systems, and reflect the practice that traditional indicators are more reliable. However, this is also reflected in new emerging indicators. Overall, both domestic and foreign evaluation systems share similar evaluation indicators. A number of indexes—annual index, the ratio of funded papers, and key retrieval systems—are included in three or more of the evaluation indicator systems.

Based on the results of the evaluation indexes of domestic and foreign journals, we summarize and refine the design principle, algorithm design, and the applicability and limitation of citation frequency, impact factor, annual index, the article funding ratio, and key retrieval system (Table 20.17).

Design Principle

Among the main indexes used to evaluate journals, the majority relate to citation rate. This stems from Garfield's citation evaluation with citation frequency and impact factor. The main evaluation index of the domestic and foreign journal evaluation system is based on the relationship among citation frequency, influence factor, and annual index. The impact range, depth, and efficiency of the journal are measured separately. The article funding index originated from China, and ISTIC was the first to adopt this index in the selection of Chinese STM journals, thus promoting the extensive influence of the article funding ratio. The article funding ratio is based on whether an article has received government funding; it is generally believed that funding indicates research trends in the field of research. To

Table 20.17 Evaluation index attributes

Property	Total citation	Impact factor	Immediacy index	Funded papers rate	Information important for the retrieval system
Design principle	Journal citation relations	Based on the citation relationship, combined with the volume of postings and the cited batches	Journal citation relations	Is the paper funded?	The information of the retrieval system
Algorithm design	In a statistical source, the total number of papers published by a particular periodical since its inception has been cited in the journal of statistical sources	The average citation rate of journal papers, that is, the ratio of citation quantity and number of journal papers	Papers published in the year of the journal that were cited	The proportion of papers funded by national, provincial level and other important funds in periodicals	Statistics on the number of times a periodical is retrieved
Applicability	The total cited frequency is applicable to the evaluation of basic research results, showing the extent to which the journal has been used and valued, and the breadth of the measures affected	The influence factor is suitable for periodical comparison in the same subject, and the influence factor is helpful to measure the periodical's knowledge communication ability and academic influence	Immediate reaction rate of periodical measure in the period of annual index	The fund thesis is advantageous to highlight the output of the project	The recognition degree and influence of periodical evaluation system by domestic and foreign periodicals
Limitations	Without considering the subject area, publication time and so on, the difference of the citation opportunity is not considered; the importance of the paper in the citation network is not considered	It is easily influenced by human factors, and the influence factors can be increased by self-citation	The measure time span is small, and can be influenced by the discipline and so on.	It is easily influenced by human factors, and the proportion cannot reflect the difference of different funds; unreasonable evaluation of nonfunded papers	The evaluation angle is single; it is not easy to quantify; it is more suitable for reference index of evaluation results
Application	Widely used, domestic and foreign periodicals to evaluate the influence factors of SCI, CSTPCD, etc.	Widely used, domestic and foreign periodicals to evaluate SCI, CSTPCD, etc. are selected cited frequency	Widely used in the evaluation of domestic and foreign periodicals by SCI, CSTPCD, etc.	Mainly used in the evaluation of Chinese periodicals CSTPCD and GCHC	Widely used, by Ei, Scopus, CSTPCD, etc.
Data sources	JCR etc.	JCR etc.	JCR etc.	CJCR etc.	webpages

study the scope or focus of research, the article funding ratio can reflect the academic quality of journals. The inclusion of a key retrieval system concerns statistics on whether the journals are included in key domestic and foreign retrieval systems. For example, if a journal is included in an authoritative and respected system such as SCI, this reflects the high academic quality and international influence of the journal.

Algorithm Design

The total citation frequency is the cumulative sum of citations and impact factor refers to a certain time span, based on either two-, three-, or five-year impact factors. That is, the annual index is cited in CSTPCD, and the funded papers and the active papers mean all papers are accounted for. It can be seen here that there are differences in the perspectives of various index measures, which can be used to evaluate the journal by the accumulated degree of impact and different time periods.

Applicability and Limitations

Regarding applicability and limitations, total citation frequency is a cumulative approach that is more appropriate for basic research subjects regarding the impact of the measurement. The timeliness of the subject is not very applicable. However, the publication of the impact factor of journals with an early publication time provides a larger window for publication, and therefore a greater chance of citation. There is less evaluation of the total citation frequency compared with journals that were published later. Garfield, based on the optimization of the total citation frequency index, stated that the influencing factors are a combination of time, the volume of journals, and citation rate. Impact factors have their limitation in that they are only applicable to the comparison and evaluation of the influence and communication ability of journals under the same subject. The annual index reflects a journal's annual efficiency, and the effect factor is only applicable to the comparison of journals within the same subject area. In view of the above index only applying to the same discipline under journal evaluation, the journal evaluation system has continued with its unceasing optimization and consummation of subject classification. Interdisciplinary journals can include a range of disciplines under the index evaluation.

Application and Data Sources

Citation frequency, impact factor, and the annual index are evaluation indexes used by JCR. These are widely used in most evaluation systems, with the continuous improvement and optimization of traditional indexes. In this chapter, we propose derivative indexes and composite indexes for evaluation, which provide a new perspective for the rational and scientific evaluation of journals. The article funding index was first used in the selection of STM journals in China. The 2001 citation report and the 2008 core journal index have increased the number of indexes they use; journals are usually included in the systems' official websites, which is mainly used in the journal evaluation system, and is an important reference index for evaluating journals.

The main indicators of Chinese and foreign evaluation systems are citation frequency, impact factor, annual index, article funding ratio, and key retrieval system. These measure the impact of the scope of the journal, depth of influence, the rate of reflection and influence, the measure of the journal, the time frame of different measures, and the difference of measurement content, and are widely applied to the main journal evaluation systems at home and abroad. Therefore, the index of citation frequency influences the journal and academic quality in the rational and scientific assessment of different dimensions.

20.5 China's STM Journals: The Development of the Boom Index and its Monitoring Function

20.5.1 Introduction to Chinese STM Journals

In 2013, a total of 9877 periodicals were published in China; that is, the average quantity of printing for each issue is 164.53 million copies, and the quantitiy of printing for total issues is 3.272 billion copies. Total quantity of impression is 19.47 billion pieces, costing 25.335 billion RMB Yuan. Compared with 2012 rate, this represents an increase of 0.1% in new titles, and the average quantity of printing for each issue droped by 1.87%. Furthermore, the quantitiy of printing for total issues showed a 2.26% reduction, the total quantity of impression was down 0.67%. However, the costs increased by 0.26%.

During the period 2009–2013, the number of Chinese journals increased slightly in 2011. However, there was a decline in the total number of journals for 2012–2013, and the average period of Chinese journals declined continuously. Regarding total prints and total sales, these showed a decrease in 2013, while the cost of journals increased.

Compared with the publication numbers of other Chinese journals, STM journals in 2009–2013 have shown periods of growth and decline (Table 20.18).

Table 20.18 Publication of Chinese journals in 2009–2013

Indicator of China's periodicals	2009	2010	2011	2012	2013
Number of natural science and technology periodicals (A)	4926	4936	4920	4953	4944
Number of periodicals (B)	9851	9884	9849	9867	9877
A/B (%)	50.01	49.94	49.95	50.20	50.06

The number of journals decreased in 2012 and 2013, whereas previously the number had increased by 0.18%. The total number of STM journals in China has accounted for about 50% of the total of journals for many years. The values for STM journals in 2013 are as follows: the total of 463.44 million copies printed with total 3.7941 billion units of printing paper. Compared with the previous year, the number of total copies printed was down 4.58%, and the number of total units of printing paper decreased by 0.01%. Thus, in 2013, while the content of Chinese STM journals remained basically unchanged, there was a decrease in the printing volume of journals and the price increased.

20.5.2 Core Chinese STM Journals

The ISTIC was set up by the Ministry of Science and Technology, and since 1987 it has engaged in the statistical analysis of Chinese STM papers. This then led to the development of the CSTPCD. The data from the database is used to annually classify and analyze the status of Chinese scientific research output. The statistical analysis results are regularly published in the public arena in the form of annual research reports and press releases. A series of research reports have been published to provide decision-making support for government departments, universities, and research institutions.

Journals of Chinese STM papers and citation databases are called Chinese STM journals. Selected Chinese STM core journals (the statistical source journals of Chinese STM papers), after rigorous peer review and quantitative evaluation, represent quality research in the field of Chinese science and technology, and they reflect the level of development in that subject area. Furthermore, it has established a dynamic exit mechanism for the selection of Chinese STM core journals. To study the scientific indexes of Chinese STM core journals, we can determine the development of Chinese STM journals, and outline the research power of various disciplines in China. The data source of this chapter's journal index is core Chinese STM journals. The 2013 CSTPCD includes the core Chinese STM journals. There were 1989 core STM journals, five less than in 2012. The number of core Chinese STM journals has declined for the second consecutive year (Table 20.19).

The distribution of the subject area of the 1989 Chinese STM journals in 2013 is as follows: industrial technology (37.31%), medical and health (33.32%), basic science category III (15.03%), agriculture, forestry, and fishery (7.83%), and other categories (6.52%). Compared with the past five years, the total number of journals has declined slightly, and the proportion of the total number of journals in the five major categories has changed little. However, compared with 2012, the proportion of basic science and agriculture, forestry and fishery has increased, the proportion of medical and health journals has declined, and the proportion of industrial technical journals has increased.

In the selection of core Chinese STM journals, the proportion of industrial technical journals and agriculture, forestry, and fishery journals is smaller than that of other journals. This indicates that the overall level of industrial technology and agriculture, forestry, and fishery journals requires some improvement, and we should pay more attention these journals in the future.

According to the statistical analysis of the 2009 Urich's International Journal Guide, the world's STM journals make up 30% of all journals, and comprehensive STM journals account for just 3%. Compared with the number of Chinese STM journals, the proportion of comprehensive journals in China is larger than that worldwide, and the proportion of medical and health journals is consistent with the world trend.

20.5.3 Citation Reports of Chinese STM Journals

In 1997, the China Scientific and Technical Papers Statistics and Analysis project group published the first *China Science and Technology Journal Citation Report*. The research group has published a new edition of the scientific and technological journal indicators. The data used in the citation reports of Chinese STM journals (core editions) are taken from ISTIC's CSTPCD. The database includes the most important scientific and technological journals in China, thus journals become statistical sources, subject to annual dynamic adjustments. By 2014, there were 1989 statistical sources of scientific papers in China. This represents the statistical analysis of the overall situation of Chinese STM papers, as well as the development of tracking research in Chinese STM journals. This has formed an annual report on China's core STM journals outlining the measurement indicators of the system. In addition, it promotes the development of Chinese STM journals and provides an evaluation basis for journal and journal management.

Table 20.19 Core Chinese STM journals 2009–2013

Year	2009	2010	2011	2012	2013
Number of science and technology core periodicals (A)	1946	1998	1998	1994	1989
Number of natural science and technology periodicals (B)	4926	4936	4920	4953	4944
A/B	39.51%	40.48%	40.61%	40.25%	40.23%

To select the core Chinese STM journals, the institute also began publishing its *Citation Report of Chinese STM Journals* (full edition) in 1998. From 2007 onwards, an expanded version of the citation report has been jointly published by ISTIC and Wanfang Data–Digital Periodicals, covering more than 6000 Chinese STM journals.

Index Analysis of Chinese STM Journals

To ensure the comprehensive, accurate, impartial and objective evaluation and use of journals, the ISTIC's citation reports are consistent with the international evaluation system. Based on the actual situation affecting Chinese journals, the *2013 Citation Report of Chinese STM Journals' Citation Report* (core edition) selected 23 measurement indicators, and the 2014 report (core edition) added a further measurement index: core open factors. These indicators basically cover and describe the various aspects of the journal. Indicators include:

(1) The journal's citation-based indicator of measurement: the core total citation frequency, core impact factor, core annual index, core citation rate, core citation number, core diffusion factor, core open factor, core authority factor, and core citation half-life.
(2) Journal source measurement index: source literature, document selection rate, reference volume, average citation rate, average number of authors, region distribution, mechanism distribution, overseas thesis ratio, article funding ratio, and reference half-life.
(3) Index of journal measurement in subject classification: total score, subject diffusion index, subject impact index, deviation ratio of the core total citation frequency, and the deviation of the core impact factor.

Among these, the journal citation measurement index mainly shows the journal's degree of use and value by readers, as well as its position and function in scientific communication, which is an important basis and objective standard to evaluate the influence of journals.

Through the statistical analysis of source literature, this chapter comprehensively describes academic level, editorial status, and degree of scientific communication of the journal, which is also an important basis for evaluating journals.

The comprehensive evaluation score is a comprehensive description of the overall position of the journal.

Table 20.20 shows the changes in the main measurement indicators for STM journals from 2001 to 2013. Since 2001, the important indicators of Chinese STM journals (except those pertaining to foreign journals) have remained unchanged. Furthermore, the values of the remaining indexes are increasing. The total citation frequency and impact factors have improved every year, of which the average citation frequency of 2011 Chinese journals finally exceeded 1000, reaching 1022 citations. This rate was 1023 in 2012 and 1180 in 2013. The 2013 rate is 5.2 times greater than the 2001 rate, and the core impact factor rose to 0.523 in 2013, twice that of 2001. These two indicators are important indicators reflecting the impact of STM journals. The annual index, that is, the citation rate of the paper in the year since 2001, has gradually increased, reaching 0.072 in 2013. The article funding ratio shows the proportion of all papers in China's core STM journals that receive state/provincial-level funding or from other key funding areas. This is also an important index to measure the academic quality of journals. From 2001 to 2013, core Chinese STM journals with funded articles show an annual increase. This rate was 0.056 in 2013. That

Table 20.20 Core Chinese STM journals: average statistics of measurement indexes

	2001	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**
Core total citations	227	278	362	434	534	650	749	804	913	971	1022	1023	1180
Core impact factor	0.264	0.294	0.348	0.386	0.407	0.444	0.469	0.445	0.452	0.463	0.454	0.493	0.523
Core immediacy index	0.045	0.048	0.056	0.053	0.052	0.055	0.054	0.055	0.057	0.06	0.059	0.068	0.072
Funded papers rate	0.34	0.36	0.38	0.41	0.45	0.47	0.46	0.46	0.49	0.51	0.53	0.53	0.56
Overseas papers[a] rate	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.02	0.02	0.023	0.02	0.02
Author number of articles	3.26	3.27	3.34	3.43	3.47	3.55	3.81	3.66	3.71	3.92	3.8	3.9	4.0
Number of references	7.36	8.21	8.81	9.27	9.91	10.55	10.01	11.96	12.64	13.41	13.97	14.85	15.9

[a] Paper with at least one author affiliated with an organization outside mainland China

is, of the 1989 core STM papers published, more than 50% of the papers are funded by the state. One of the indicators showing the international level of the journals, the rate of overseas papers, remained the same, staying between 0.01–0.02 (0.01 in 2007 and 2008). The average number of authors and citations have increased annually, from 3.26 and 7.36 in 2001 to 4.0 and 15.9 in 2013, respectively.

Regarding core citation rates, these show almost linear growth by 2011. The 2012 core average citation rate represents a significant slowdown, almost the same as the 2011 rate. However, in 2013, the core average citation rate increased to 1180. Table 20.20 shows the average core impact factor and the average core annual index for 2001–2013. The growth in the average core impact factor peaked in 2007, and after five years of decline, it increased again in 2012–2013, higher than the 2007 rates (0.493 and 0.523). The average core impact factor for 2001–2013 shows a steady upward trend, peaking in 2003. This was followed by a five-year decline, a three-year fine-tuning period, and then an increase in 2012–2013 to a high of 0.072. This shows that despite the fluctuations in the citations of the core Chinese STM journals, they now enjoy a steady increase.

At the same time, the average number of core Chinese STM journals has continued to increase, while the growth of average core influence factor in 2001–2013 has slowed. The core average of total citation frequency and the core average impact factor peaked in 2003, slowing from 2004 onwards. The growth of the average total citation frequency between 2004 and 2013 experienced three troughs, in 2004, 2008 and 2012. Interestingly, each trough is followed by three years of growth, but then falls again. The growth rate during the nadir period shows a gradual decrease. The 2012 growth rate is almost 0, while the growth rate of the core journals in 2013 increased again, reaching 0.153 for the first time in nearly four years. The average core impact factor experienced three troughs between 2004 and 2013 (i. e., in 2005, 2008 and 2011), and the growth rates of the average core factors in 2008 and 2011 were 0.05 and 0.02 respectively. The 2008 rate is the lowest for the period (−0.05), the average core impact factor did not increase, and the average core factor increased in 2012–2013, with a growth rate of 0.086 in 2012.

According to the index analysis of papers published by STM journals, the number of key funding grants and papers published in STM journals reflects the academic quality and the level of journals. For academic journals especially, this index appears to be very important. The publishing of foreign articles is an important index of a journal's level of internationalization. The article funding ratio increased from 0.34 in 2001 to 0.47 in 2006, and then experienced a slight decline for 2007–2008. It enjoyed an increasing trend during 2009–2011 reaching 0.56 in 2013. Thus, more than half of the papers published are supported by funding or grants from provincial and ministerial-level government departments.

In recent years, China has worked hard to increase its investment in scientific research. The success of the 11th Five-year Plan has meant that a large number of scientific research projects have produced many scientific papers. Overseas papers have remained at 2% since 2001–2006, falling to 1% in 2007–2008, and climbing to 2% in 2013. This shows that the international volume of Chinese STM core journals has been hovering at 1–2%, and therefore Chinese journals have a low level of internationalization.

The average citation index refers to the average reference quantity of each paper in a journal. It is a relative index to measure the degree of scientific communication of STM journals and the ability to absorb external information. Furthermore, the standardization of reference documents is also an important index that reflects the standardization degree of Chinese academic journals and the integration of international scientific research. The average citation number of core Chinese STM journals increased between 2001 and 2013, with a slight decrease in 2007 (however, 2007 also represented the first time the rate rose above 10, reaching 10.01). The average citation rate in 2013 was 15.9, 2.16 times that of the 2001 rate.

Over the past 20 years, there has been a noticeable development within the statistical and analytical research contained in Chinese STM papers. Furthermore, with the extensive publicity of the evaluation system of STM journals and with more and more Chinese and international researchers, researchers now pay greater attention to the completeness and standardization of articles. Aware of the importance of bibliographic references, a vast number of STM journal editors are acknowledging that the preservation of the objective of reference documents is an important channel for journals and academic exchanges. Therefore, the average number of citations has gradually improved. From 2001–2012, the average number of authors in Chinese STM journals hovered between 3.26 and 3.92. In 2013, the average number of authors reached 4.

The Status of Core Chinese STM Journals

In 2013, there were a total of 1989 core Chinese STM journals. They published 502 393 papers, a 1.49% decrease since 2012. On average, for each core journal, there are 252.59 source articles.

The quantity of source literature, that is, the quantity of the journals, refers to the amount of information contained in the journal, specifically the number of papers published in a journal year. It should be noted that Chinese STM papers and citation databases in the collection of papers refers to the choice of journal papers. We refer to the volume of papers that are in academic journals of scientific papers and research bulletins. These present research findings on new technologies, materials, processes and new products, and include research papers on basic medical theories in medical journals and important clinical practice summary reports as well as literature reviews.

In 2013, there were 623 journals that included more sources of literature than the average number of Chinese journals (in 2012 there were just eight). There were five journals with more than 2000 documents, with one including 2367 articles. There were journals with more than 1000 papers.

Between 2004 and 2013, the propotion of the journals whose number of publishing paper each year is no more than 50 showed a steady decline. Papers published in 100–200 journals accounted for the highest proportion. These show an overall slight downward trend. In ten years of core Chinese STM journals, 40% of all journals were included in 100–200 journals; the proportion of published articles in 50–100 journals has declined since 2004. The proportion of other journals is increasing, which shows that the information capacity of China's core STM journals is expanding and the volume of journals is increasing.

In this chapter, we also use statistics to determine the relation between subject classification and the region of article numbers of a journal. The percentage of basic subjects journals is 61.3% in regions where the volume of article is less than or equal to 50, which is much higher than that of the other four categories (engineering, agriculture, medicine, others). With an increasing volume of papers, the proportion of basic subject journals declined sharply, and the proportion of basic subject journals fell to 5.76% in regions with a volume of more than 500. Regarding the category of agriculture, forestry fishery, and husbandry compared with other three subject clusters (basic science categories, medical and health categories, industrial technology categories), the proportion of journals varies little in each region where the amount of papers is distributed. Journals in the medical and health categories show a decrease in volume and the proportion of journals obviously decreased. In regions where the volume is less than 300, the proportion of journals shows a near-linear decrease. Industrial technology journals in regions with a volume greater than 50 have a small distribution, between 36.55% and 39%, and the proportion of journals in regions less than 50 is sharply reduced to 18.75%. This shows that industrial technology and medical and health journals are distributed in larger areas, and basic science journals are distributed in more regions but in smaller volumes.

Subject Analysis of Core Chinese STM Journals

Compared with the previous version, the *2013 Citation Report of Chinese STM Journals* (core edition), shows considerable changes to the journal subject classification. The 2013 version is the latest implementation of the Discipline Classification and Code (national standard Gb 113745). We will use core Chinese STM journals to redefine the subject. The original 61 disciplines have been expanded to 113 disciplines. The 2014 edition of the citation report was further adjusted, adding two new subjects. Furthermore, the subject classification of journals was adjusted and divided, and interdisciplinary journals were classified as two or three subjects. The new subject classification system embodies the development and evolution of scientific research disciplines, which is more in line with the overall situation of the development of science and technology in China and the actual distribution of Chinese STM journals. In 2013, 1989 core Chinese STM journals covered various disciplines. *The Journal of Engineering Technology University*, the *University Journal* and the *Journal of Medical Universities* occupy the top three rankings, and the *Journal of Engineering and Technology University* is 108th. In all core journals, as the top proportion part, there are 13.68% of them are the academic transactions or journals of universities (including comprehensive universities, polytechnic universities and medical universities). For example there are 108 journals of polytechnic universites in all 1989 core journals. As a China characteristics, big number of journal of universities is a major force of Chinese STM journals.

The average impact factor and citation frequency of core Chinese STM journals in 2013 was 0.523 and 1180, respectively. A total of 58 subjects have impact factors higher than the average, representing 11 more than in 2012. Furthermore, there are 52 subjects with higher-than-average citation frequencies (ten fewer than in 2012). The top three impact factors were in grassland, atmospheric science, and geography, and the top three citation frequencies were ecology, nursing, and civil engineering. The influence factor has a strong correlation with the subject field, and the impact factors of different disciplines show clear differences. Because of the significant difference among subjects, we focus on the subjects' 2013 values.

For all 115 subjects in 2013, the midian value of total citations is more than 1000 times in 34 subjects. Among these, medical and health journals accounted for 50%, industrial technical journals accounted for 20.59%, basic science journals accounted for 17.65%, and agriculture, forestry, and fishery journals accounted for 8.82%. The top three subjects were ecology, nursing, and ecology. Geography, atmospheric science, and ecology are among the higher values of the 2013 discipline factors. Subjects with lower impact factors include comprehensive, metallurgical technology, and mathematics. Therefore, it is necessary to compare the impact factors of STM journals with the average level of the subject.

A Regional Analysis of Core Chinese STM Journals

The regional distribution number refers to the number of regions covered by the author of the source journal, which is calculated according to China's 31 provinces (municipalities).

Generally speaking, the geographical distribution of a journal can be used to determine whether the journal is a more widely covered journal, its influence in the whole country, and whether the regional distribution covers more than 20 provinces (cities). Journals meeting these criteria are considered national journals.

After 2004, the number of journals in China's core STM journals was more than or equal to the 30 provinces (cities), with 2011 having the largest number of journals, 106 in 2012, and an increase of 5% in 2013.

The regional distribution of more than 20 provinces continues to increase, with more than 64.23% of all journals. That is, in 2013, more than 64% of the journals are national STM journals. Moreover, the proportion of journals with a regional distribution of less than ten continues to decline.

Publishing Cycle of Core Chinese STM Journals

Because the speed of publication is an important basis of scientific discovery, the shorter the journal publication period, the stronger the ability to attract excellent manuscripts, and the higher the impact factor. Research shows that the publishing cycle of China's STM journals is reducing annually.

In 2013, the publishing cycle of China's core STM journals was further reduced, and monthly publications rose from 28.73% in 2007 to 37.45% in 2013. Additionally, bimonthly publications decreased from 52.49% in 2007 to 48.82% in 2013, as more bimonthly journals became monthly journals. Quarterly journals fell from 13.22% in 2008 to 9.6% in 2013. Compared with journal publication periods in 2012, the number of journals remains basically unchanged in all other periods. While the overall publication cycle of STM journals has been shortened, close to 50% of journals are published bimonthly.

From a classification perspective, basic science journals and those on agriculture, forestry, and fishery and comprehensive journals have relatively long publishing cycles, and quarterly and bimonthly ratios account for about 70%, which shows that in these three categories, journals are mainly published as quarterly or bimonthly journals. The publication cycle of basic science journals is:

- Quarterly: 22.54%
- Bimonthly: 49.21%
- Monthly: 26.98%
- Others: 1.27%.

For industrial and technical journals: 58.18% are published as either quarterly or bimonthly journals, up from 57.83% in 2012. In other words, more than 50% of the journals in industrial technical journals are quarterly and bimonthly, and this is 10% higher than the proportion of basic science, agroforestry, and other journals. The distribution for medical and health journals is: 46.63% of journals in this field are published as quarterly and bimonthly journals, and 53.37% are published monthly. Thus, more than 50% of the medical and health journals are monthly. 2012 data does not contain information on fortnightly and weekly publications, so no comparison can be made.

The world's largest journal directory guide, the Ulrich Journal Guide, surveyed 50 443 academic journals and found that quarterly journals accounted for nearly 30% of the total number of academic journals. Furthermore, quarterly, biannual, bimonthly, annual, and monthly publications accounted for 80.9% of the five publication cycles. Thus, for academic journals around the world, the quarterly journal is the most important publication cycle, accounting for 29.5% of all journals; this is followed by semiannual, bimonthly, and annual editions. Unlike in China, bimonthly publication is not the most common format.

In 2013, the JCR included a total of 8194 journals, with a diverse range of publication cycles. There were 37 types of publication cycle, with a maximum number is 60 issues per year and a minimum number is 1 issue per year. To facilitate our comparison with Chinese journals, we will include JCR journals according to the publication period, reducing the number of categories. We will merge weekly and more than weekly journals as a weekly comparison with Chinese journals. For example, 26 kinds of JCR journals are published weekly, whereas there is only one weekly journal in China. Sixty are published annually. The weekly journal and the 26 other more than weekly journals are merged into 27 weekly journals to aid in our comparison with Chinese journals.

It is clear that the 2013 JCR includes the largest proportion bimonthly journals (33.72%), followed by quarterly (28.08%), monthly (27.29%), and then triannual, semiannual, and annual (6.73%). Regarding China's

2013 core STM journals, the largest percentage are bimonthly publications. The JCR includes bimonthly, quarterly, triannual, and annual publications (68.53%). China's core STM journals published bimonthly and quarterly make up 58.43% of such journals (there are no triannual, semiannual, and annual journals). Therefore, the publication cycle of core Chinese STM journals is shorter than that of the journals in the JCR.

The 2013 content of SCI includes the distribution of 139 Chinese STM journals. The publication of journals has also diversified. In contrast to 2012, the proportion of monthly journals in 2013 (35.97%) exceeds bimonthly publications (34.53%). These two formats are followed by quarterly journals (26.62%), which has declined since 2012. Despite the increase in the monthly ratio, the ratio of bimonthly and quarterly journals is still high, at 61.15%, in line with the publication cycle of journals in the JCR.

20.5.4 An International Comparison of Chinese STM Journals

We now compare the average citation frequency, average impact factor, and average annual index of core Chinese STM journals and JCR journals for 2011–2012. The values for these three indicators have increased for JCR journals, which is consistent with the growth of China's core STM journals. However, our absolute value is not at a comparable level to international journals. International journals enjoy much higher scores than core Chinese STM journals. The average citation frequency and the average impact factor are four times higher than the annual index, which is six times higher.

In the issue of number of articles, China's core STM journals and journals covered by JCR are very different. In the region of volume of articles of more than 100, the percentage of core Chinese STM journals is obviously higher than JCR journals. The percentage of China's core STM journals are significantly lower than that of JCR journals when we look at the proportion of the journal in areas where number of papers is less than 100. Regarding areas where the paper number is less than 50, the gap is particularly obvious. China's core STM journals account for 1.61%, and JCR journals account for 33.35%, 31.96%, and 32.26%. China's core STM journals published more than 100 papers, accounting for more than 82% of the total number of journals. The total number of JCR journals publishing more than 100 papers is less than 44%. The number of journals with less than 100 papers accounts for more than 60% of the total number of journals. This shows that the number of papers published in Chinese STM core journals is higher than the number of papers published by JCR.

In 2013, the SCI database included 139 Chinese journals (Table 20.21). The main evaluation indexes of JCR include total citations, impact factor, immediacy index, current number of papers, and article half-life (citation life). One of the most important indicators is the ranking of the journal in this discipline according to its impact factor (journal rank in categories). The impact factor is used to determine whether a journal is considered a Q1 journal. A Q1 award shows that the journal is a top-ranked journal in that particular discipline and it can also be called the top academic journal in that subject area. Based on the impact factor, a journal will be either a Q1, Q2, Q3, or Q4 journal.

The influence factors of JCR-selected Chinese journals in 2013 concern various subjects. A total of eight journals gained a Q1 ranking (compared with seven in 2012), and 34 are Q2 journals (six in 2012). There were 42 Chinese STM journals in 2013 ranked in the middle of the subject (and only nine in 2012). According to data from 2009 and 2011–2013, we note a gradual increase in the number of Chinese journals entering their corresponding disciplines.

In 2013, the following retrieval systems included Chinese STM journals: SCI, 139 journals; Ei, 216; MEDLINE, 106; SSCI, 2, and Scopus, 776.

The quality of Chinese STM journals has also undergone a process of development and change. In 1987, SCI selected only 11 Chinese journals, accounting for 0.3% of journals worldwide, and Ei selected 20 Chinese journals. For more than 20 years, the ranking of Chinese STM journals has been improving, and their influence in the world retrieval system has become greater. The number of China's STM journals is slowly growing, and have now passed the stage of quality improvement. Our aim is for these journals to move towards a stage of comprehensive revitalization.

20.5.5 Development Survey of Chinese STM Journals

Development Policy of of STM Journals

The relevant policies of management departments at all levels regarding STM periodicals play an important role in the development of these journals. A question-

Table 20.21 The number of Chinese STM journals included in SCI and Ei 2002–2013

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
SCI	69	78	78	78	78	104	108	115	128	134	135	139
Ei	108	119	152	141	163	174	197	217	210	211	207	216

naire on impact factors and the development of Chinese STM journals has investigated concerns relating to the development policy of STM journals. Of the 541 questionnaires collected, 66.17% of respondents expressed great concern, 26.5% expressed concern, 5.5% said they were aware, and only 2.03% answered that they were not sure. The investigation shows that the relevant personnel of STM journals pay attention to "the relevant policies of the development of STM journals".

In the survey respondents were given the following statement: "The existing policy adapts to the development of your journals". A total of 348 (64.33%) respondents stated "basic adaptation" had occurred, indicating that the majority of respondents affirmed the policy of STM journals. In contrast, 193 respondents selected "further promotion required", accounting for 35.67%, indicating that the relevant policies for the development of STM journals still need to be improved and upgraded. Finally, 193 respondents answered that the existing policies have not adapted to the development of STM journals' specific performance. This mainly relates to evaluation system, management system, system reform, policy guidance, and support efforts. Additionally, 74 respondents mentioned the issue of scientific evaluation, accounting for 193 responses of "concrete performance" (38.34%). This largely refers to the establishment of the evaluation of STM journal targets, a STM journal evaluation body to address negative journal development, as well as scientific research personnel, organizations, research results, and so on. The contrary of direction between appraisal targets cause the disadvantageous influence for STM journal development. Forty-one respondents (21.24%) regarded the specific performance of STM journals as a "concrete performance". This includes the strengthening of funds, manpower, technology and other support issues. It also refers to the support for Chinese journals, English-language journals and others. Support also needs to come from the director, as well as sponsorship support. Thirty-three (17.09%) respondents answered "concrete performance" regarding reform efforts. This includes the specific issues of autonomy and marketization. Furthermore, 23 (11.91%) respondents answered "concrete performance" in the management of STM journals. This referred to the issue of examination and approval, academic journal identification, and other specific issues. Seventeen (8.9%) respondents also stated "concrete performance" regarding the unclear policy orientation of STM journals. A further five respondents mentioned the quality of science and technology journals, new media applications, digital issues, and other specific problems.

In recent years, the State Administration of Radio, Film, and Television (SARFT), the Ministry of Science and Technology, CAS, and other competent departments involved in STM journals have proposed a number of policies to promote STM journal development. In 2000, the Ministry of Science and Technology, through its *Strategic Study of China's Fine Science and Technology Journals* and *China through the Service and Guarantee System of Fine STM Journals*, proposed the concept of *Quality Chinese STM Journals* and sought to promote the construction of quality STM journals. In 2011, the Central Committee of the Communist Party of China and the State Council aimed to "deepen the reform of the nonpolitical publishing system", and implement the reform of the publishing system governing nonpolitical newspapers and journals. In 2013, the Chinese Association of Science and Technology, Ministry of Finance, Ministry of Education, the General Administration of Press and Publication, SARFT, CAS, and the China Institute of Engineering made efforts to implement the *Program to Promote the International Influence of China's Scientific and Technological Journals*. The aim was to promote the internationalization of STM journals. This is the largest and most far-reaching financial aid project in this area to date. On 14 April 2014, SARFT reported intentions to standardize the publishing order of academic journals to promote their healthy development. On 11 September 2014, plans were announced to carry out the accreditation of academic journal qualifications, and on 18 November, the first list of identified academic journals was publicized. On 15 May 2014, CAS and the Natural Science Foundation issued a policy statement: public funds would be made available for scientific research projects to promote open access to scientific research. On 18 August 18 2014, at the fourth meeting on the comprehensive and deepening reform, the Central Committee considered the adoption of guidance regarding the promotion of the integration of traditional media and emerging media, and the wave of convergence between traditional media and new media. This was an issue of general concern and received many responses from STM journals.

In the questionnaire survey on impact factors in the development of Chinese STM journals, 541 respondents responded to the statement about "the influence of the existing STM journal policy on the impact of your journal". Of the six listed responses, the respondents answered as follows: speeding up the integration of new media, 52.68%; the transformation of nonpolitical journals into enterprises, 29.94%; the examination and identification of academic journals, 41.96%; the promotion of international influence of journals, 54.71%; the construction of quality journals, 57.86%; and the proportion of open acquisition of scientific papers funded by public funds, 27.36%. In addition, 24 respondents suggested other specific policies, which relate to the evaluation of STM journals and their impact.

The questionnaire also investigated the role of the relevant policies and measures of management at all levels. According to the results for multiple selection, 409 (75.6%) respondents considered that the relevant policies and measures of the organization level had a great effect on the development of STM journals. Furthermore, 344 (63.59%) and 343 (63.4%) respondents selected "national and local press and publication management departments" and "the competent departments and units of journals", respectively, for the development of STM journals. A total of 172 respondents considered that the relevant policies of "social groups such as society" had a great effect on the development of STM journals.

The Competitive Environment of Chinese STM Journals

The questionnaire survey also investigated the competitive environment of Chinese STM journals. When replying to a question concerning "the number of similar journals in your subject", 60.44% of respondents considered that there was an "appropriate" number, indicating that the subject distribution of Chinese STM journals is largely reasonable. At the same time, it should be noted that 31.05% believe that the number of similar journals in the subject is too large, which indicates that the homogeneity of journals in some subjects is a serious problem, raising issues of strong fierce competition and an oversupply of resources. In addition, 8.5% of respondents thought that there were "too few" journals in the same subject area, indicating that there is only a small supply of resources in that discipline. Furthermore, in response to a question on the "competitive strength between your journal and the same subject", 112 (20.7%) replied "very strong", 253 (46.77%) answered "intense", and 166 (30.68%) answered "average". The results show that there is strong competition among Chinese STM journals.

The questionnaire also investigated the main aspects of competition among Chinese STM journals. According to the survey results, 95.19% of the respondents stated there was competition regarding "author and manuscript". At the same time, competition between Chinese STM journals and other journals was reflected in readership, funding, reporting speed, reviewers, and editors (34.38%, 29.57%, 24.77, and 20.33%, respectively).

When responding to a question about the "management policy of the same kind of journal", 416 (76.89%) respondents indicated that "guiding the construction of characteristic journals and reducing the degree of homogeneity" were important. A total of 348 (64.33%) respondents considered that "controlling quantity, supervising quality, and maintaining a healthy competitive environment" were effective policies. A total of 276 (51.02%) and 267 (49.35%) respondents selected "focus on helping the strong colleagues to eliminate low-level duplication of the journal" and "decentralization, encourage the innovation of scientific and technological journals", respectively. These results are very similar.

20.5.6 Development of the Chinese Science and Technology Boom Index

The Compilation of the Boom Index of STM Journals

The boom index of STM journals refers to the creation of a synthetic index (CI), the index of a comprehensive STM journal industry that includes demand, supply, policy, cost, and income. These features make up the boom index of STM journals. The boom index of STM journals is a dynamic and comprehensive industry index that can reflect the prosperity of the industry. It also monitors and predicts fluctuations in the STM journal market.

The index compilation method adopted by the industrial boom index is a synthetic index method used by the US Department of Commerce. The National Housing Boom Index is another method that can be used to calculate a synthetic index. The China Purchasing Manager index adopts the diffusion index method, and the development of the Yangtze River Industry index is a comprehensive use of the boom questionnaire and diffusion index.

In this chapter, using the above compiling method for the boom index, we use the synthesis index compiling method of the US Department of Commerce, which will reflect the health of the STM journal industry. The boom indicator system for STM journals includes a synthesis index, a consistent synthesis index, and a lag synthetic index. These will identify any recessions, recovery, expansions, and contractions within STM journals. Furthermore, the boom index has monitoring and forecasting functions.

The Classification of a Prosperity Index for STM Journals

Advance Index (Leading Indicators). The first index refers to the index of peaks or valleys before the macroeconomic fluctuation reaches the peak or trough. The general antecedent index should meet the following conditions: The peak (or valley) point of each special cycle of the sequence is at least three months prior to the reference cycle. This first relationship is stable; that is, there is not much difference between the super prophase and the sequence in the last two cycles. The peak (or valley) of the special cycle should remain ahead, and the lead time is more than three months. The economic nature of the index has a definite and clear antecedent. The first index has an antecedent character. There are are two reasons for this. First, some advance indicators act as forecast indicators of business trends, such as "predicting business conditions". Second, future economic activities will result in business changes; for example, such indicators include hiring new staff, new construction areas, and new residential areas. In many countries, the advance index is regarded as an important basis for short-term forecasts.

There are four leading indexes for STM journals: economic index, policy index, academic environment index, and technical progress index. Scientific research and experimental development funds can be used as an index to evaluate the input, scale, and intensity of science and technology in a country or region. These will reflect the future economic index of the STM journal industry. The policy index mainly determines which policy will be applied to the STM journal industry; for example, the journal quality project and the project for journals with international influence.

Consistent Indicator (Coincident Indicators). The consensus index (also known as the synchronous indicator) of STM journals measures the peaks and troughs and the date of the benchmark economic cycle fluctuations. Therefore, the consistent indicators reflect the current situation of the boom.

The consistent index of STM journals mainly includes:

(i) Sequence overall operation level: the publication scale of journals in the news publishing industry
(ii) Journal group level: based on the point of view of citation networks, the paper extracts the monitoring index of journal group level from the aspects of journal subject, publishing language, and adopting new media technology
(iii) Journal level: reflecting the operation index of STMjournal industry from individual journal level
(iv) Thesis level: to reflect the operating index of STM journal industry from a single thesis level.

Lagging Indicator (Lagging Indicators). The lagging indicator refers to the benchmark turning point, lagging behind the economic cycle fluctuations. The effect of the lag index is that its peak or valley appearance can confirm that the peaks or troughs of the economic fluctuations are indeed appearing.

A lagging index of STM journals includes the following features: qualitative index: a qualitative index that can be used to measure the development level of STM journals, including the status of major international retrieval systems; quantitative index: including the average citation number and reference number of journals; and the liquidity index: refers to the flow of the pyramid structure in the journal hierarchy. At the top of the journal pyramid is the journal sequence of TOP100, followed by quality STM journals, core STM core journals, and general journals.

A lagging index of STM journals includes the following features:

(i) Qualitative index: a qualitative index that can be used to measure the development level of STM journals, including the status of major international retrieval systems
(ii) Quantitative index: including the average citation number and reference number of journals
(iii) And the liquidity index: refers to the flow of the pyramid structure in the journal hierarchy.

The Intrinsic Mechanism of the Consistent Prosperity Index

The consistent prosperity index of STM journals is derived from the statistic index of the STM journal industry. This index of prosperity is compiled by the consistent index of prosperity. According to the selection principle of the boom index and the compilation methods of other boom indexes, the index of the industry prosperity of STM journals is composed of demand, supply, policy, labor, cost, and income.

(i) Production index of STM journals: publication quantity of STM journals; policy indicators of STM journals: quality journals, international influence, funded journals
(ii) Input index of STM journals: scientific research and experimental development outlay
(iii) And employment index of STM journals: number of editorial staff and number of journals. All indexes of STM journals are monthly or quarterly indicators, and the monthly indices are more timely and suitable for use in the boom index. However, statistical indexes are very difficult to obtain for STM journals, and the monthly data collection is very important. Therefore, the index of STM journals uses annual STM journals to meet the requirements.

Indicator System of Boom Index of STM Journals

Principles of Construction. The evaluation index concerns certain characteristic of the evaluation object and its quantitative performance, which not only identifies particular characteristics of the evaluation object (i. e., nature), but also the quantity of the evaluation object and the dual function of qualitative and quantitative cognition. According to the needs of the evaluation task and target, the evaluation indicator system, which comprehensively and systematically reflects a series of evaluation objects, has a relatively complete and structured relationship. The evaluation index and indicator system is a reflection of the whole or part of the subject, and the evaluation index and indicator system accurately reflect the true degree of certain matters. This is the basic guarantee of scientific and technological evaluation results.

The formation of an evaluation indicator system is a complex process because the indicator system itself is a complex system. Furthermore, it is an organic system composed of a series of interrelated evaluation indexes. Therefore, the design of a complete, scientific, and systematic indicator system is not a simple and random process, but a complex process of multiple interconnected features. The SMART (SMART) system is used by the World Bank and many national government departments and organizations as a guideline for the design of evaluation indicators. The SMART criteria essentially provides a description of the basic requirements for general evaluation index design.

It is a complicated task to comprehensively evaluate the prosperity index of STM journals. Therefore, the establishment of an indicator system of STM journals' prosperity index should follow the following principles.

Principle of Purpose. The selected index should be able to objectively describe the essential characteristics of the object and should serve the purpose of the evaluation. The construction of the boom index of STM journals should take any development trends and impact factors of STM journals into consideration.

Principles of Science. This is the basis to ensure that the evaluation structure is accurate and reasonable. That is, the meaning of the indicator system should be accurate and clear. The creation of an evaluation indicator system must be complete, so it can fully reflect the essential characteristics of the evaluation object. There must be a logical relationship between the indicators in the system, and the indexes should be adapted to the evaluation object and the evaluation target. In this way, the characteristics of the boom index of STM journals are reflected from different perspectives. Furthermore, the definition of the index should be clear and exact and the method needs to be scientific to ensure the credibility of the index.

Principles of Operability. All indicators should be able to reasonably quantify and have comparability and realistic feasibility.

The Principle of Completeness. The indexes in the evaluation indicator system must be fully understood and the ability of each cannot be evaluated in isolation. We need to focus on the characteristics of the boom index of STM journals and comprehensively evaluate the index of STM journals in different levels and to different degrees. To fully evaluate the prosperity index of STM journals, the indicator system should be comprehensive and objective, and able to reflect the overall situation of the evaluation object. The design of scientific and technical personnel evaluation indicators need to reflect the requirements of the introduction of scientific and technological journals. Furthermore, the evaluation indicator system should be able to comprehensively reflect the introduction of STM journals and evaluation objectives. The objective elements require comprehensive consideration to achieve the overall goal. Of course, the system should be kept as simple as possible, and this is conducive to the evaluation of the development and evaluation of the reliability of the improvement.

Construction Method. Generally speaking, the methods of scientific evaluation indicator system usually include the following processes:

A. *Evaluation indicator system-building methodology (primary method)*
   The construction of an evaluation indicator system is mainly based on questionnaire surveys, AHP, frequency statistical method, theoretical analysis method, and an expert investigation method (e.g., the Delphi method); thus, the primary indicator system is formed.
B. *Evaluation indicator system test method (preferred method)*
   The indicator system test mainly uses various qualitative and quantitative methods to detect the integrity, systematicness, accuracy, feasibility, reliability, scientificity, relevance, coordination, and redundancy of the indicator system. The method of expert judgment is generally based on quantitative testing.
C. *Evaluation indicator optimization methodology*
   The optimization of the evaluation indicator system structure mainly concerns the depth of the levels, the number of A levels, and the existence of a network structure. It can also be a combination of qualitative and quantitative analysis.
D. *Quantification and treatment of indicators of evaluation*
   Index quantification (i.e., the determination of the index attribute value) usually concerns the quantification of a quantitative index and a qualitative

index. Quantitative indicators are generally quantified and investigated. However, qualitative index quantification can be divided into two kinds: a direct quantification method and an indirect quantification method according to the features of the object. The direct quantification method gives a quantitative quantity (such as the direct scoring method) and a certain quality mark line. The indirect quantification law represents a set of all the possible values of the qualitative index. The value of each unit is converted to a quantity by registering the values of the variable and then quantifying the elements in the *qualitative index set* to that value.

Because the evaluation indicator system reflects the scale and level of the specific evaluation object from multiple perspectives and levels, the evaluation indicator system is an information system that reflects the subject of the evaluation. It also constructs an information system that reflects the whole picture or the important characteristics of the subject. The structure of the system generally includes the configuration of system elements and the arrangement of the system structure. In an evaluation indicator system, each index is the element of the system, and the interrelation of each index is the system structure. An important feature of the system is hierarchy; therefore, in the construction of a STM evaluation indicator system, AHP is generally used to establish the hierarchical structure model of the indicator system. Then, the index is used to filter and optimize the structure of the indicator system.

AHP is a practical multiattribute evaluation method developed by Professor Saaty in 1971. AHP is a combination of qualitative and quantitative analysis, and integrates qualitative and quantitative analysis. It is a simulation of human decision-making processes and can be used to solve complex multifactor systems, especially difficult-to-quantify social systems. Since 1982, the following processes have been applied in China: energy policy analysis, industrial structure research, STM achievement evaluation, development strategy planning, and personnel assessment.

Relying on qualitative analysis and decision-making based on experience, the results often lead to decision-making errors. A large number of mathematical methods rely solely on mathematical models to solve decision-making problems, which places an uneven emphasis on quantitative analysis. For many economic and social problems, it is difficult to rely solely on quantitative mathematical models. Although theoretically speaking a mathematical model seems appropriate, it is not always suitable for decision-making and the prediction of practical activities. Therefore, quantitative mathematical methods are not omnipresent. In addition, there are a number of factors that cannot be quantified in the decision-making process. It is necessary to rely on the experience and knowledge of the decision-maker to make a judgment even if the model is able to quantitatively describe the factors, in order to rationally use quantitative technology and establish the correct mathematical model. Furthermore, the analyst's personal experience and judgment cannot be completely removed and affect the whole process of decision-making. AHP is able to decompose the elements of decision-making into a target, criterion, scheme and so on, on the basis of which the author provides a quantitative description of human subjective judgment.

AHP decomposes a complex problem into several levels, and establishes an ordered hierarchy (i. e., a hierarchical structure model) in which the elements in each level have roughly equal status and have a certain connection with the previous level and the next level. In this way, people's experience and judgment can be expressed and dealt with as a quantity. That is, when comparing the importance of 22 factors, we use the results of 22 comparisons as the elements of the judgment matrix. To obtain each factor's relative importance in the sorting results, we construct the judgment matrix by solving the matrix of the largest feature roots and corresponding eigenvectors. A rigorous logical analysis and statistical test of the results of the comparisons and judgments ensure that the judgment elements and judgment matrices are consistent in the synthesis process.

The main steps of the AHP approach include the following aspects: structuring and systematizing the assessment of the problem, listing the related attributes, and establishing the target hierarchy. The goal-level architecture of AHP can make use of group decision-making or expert assistance, such as Delphi, brainstorming, or nominal group technology. The set of attributes within each level of the target hierarchy is based on the previous level of the target, and by repeatedly evaluating and correcting the selected attributes, the principles of integrity, deconstruction, scalability, repeatability, and minimization are ensured. The evaluation attribute and the paired comparison matrix of different schemes are established under each attribute. After constructing the target hierarchy, we will evaluate the relative importance of these two attributes to the upper level target, and then target any two of the attributes at a certain level. This method enables policymakers to focus on the judgments, and through the analysis of architecture and patterns, they will not lose control of the problem. The relative weights of each attribute and the relative evaluation values of each scheme are computed using eigenvectors. Consistency can be verified either by correcting the consistency requirement or by reopening the comparison matrix for the corresponding attribute.

Construction of the Boom Index of STM Journals

Investigation into the Influential Factors of the Development of Chinese STM Journals. To obtain more accurate and comprehensive indexes, we use surveys to investigate those factors that influence the development of Chinese STM journals. A total of 541 questionnaires were collected.

From the survey results, the existing policy of STM journals has a considerable influence on the construction of quality journals and the promotion of the plan to increase the international influence of Chinese journals. Another important factor is the acceleration of and integration with new media.

When questioned about the most important development aspect, 85% of respondents stated sufficient funding, while 80% responded stable staffing. Furthermore, 54.9% of the respondents thought that STM journals now enjoy a stable development pattern and that the development environment has little influence.

Constructing a Hierarchical Framework for the Boom Index of STM Journals. Based on the questionnaire analysis, this chapter identifies the boom index of STM journals and modifies and perfects the established evaluation indicator system using the expert consultation method. This mainly revolves around whether the evaluation index is reasonable, whether it can fully reflect the introduction of the boom index, and whether it is convincing enough to obtain the opinions of experts.

According to revised views to improve the evaluation framework and indicators system, the experts' feedback will be used to modify the index. After repeated discussions, the final formulation of a complete scientific and technological journals indicator system using a hierarchical structure model is described below.

Indicator description:

- III-1 Scientific research and experimental development funding: refers to the research and development funds provided by national science and technology funding into statistical journals.
- III-2 Chinese Journal Support Fund: selected by the Chinese Science and Technology Association led by the quality of scientific and technical journals to support the project.
- III-3 English Journal Support Fund: the selection of internationally influential journals funded by CAS.
- III-4 The number of new journals: the number of newly established journals approved by GapP each year.
- III-5 New financial support: refers to the increase of national financial science and technology allocations.
- III-6 Total number of journals published: the total number of STM journals published by the General Administration of Press and Publication.
- III-7 Average period in print: the average period of the annual publishing by the General Administration of Press and Publication.
- III-8 Citation frequency: the total average citation frequency of STM core journals.
- III-9 Impact factor: the average impact factor of STM core journals.
- III-10 Annual index: the average annual index of STM core journals.
- III-11 Literature sources: sources of literature.
- III-12 Size of editorial team: number of registered professional editors nationwide; national publishing industry practitioners.
- III-13 Citation average: the average number of citations per STM journal.
- III-14 Article funding ratio: the proportion of provincial and ministerial-funded papers in scientific journals.
- III-15 SCI journals: the number of Chinese journals included in the SCI database.
- III-16 Ei journals: the number of Chinese journals included in the Ei database.
- III-17 MEDLINE journals: the number of Chinese mainland journals included in the MEDLINE database.

The Determination of the Index Weight of the Boom Index of STM Journals. The basic principle of AHP is the principle of sequencing, which is the basis of decision-making. The goal and characteristics of the boom index of STM journals, and the comparison of each level of attributes, constructs a paired comparison matrix between the attributes of the targets. According to the construction of the judgment matrix, and the consistency of the test, the index weights can be determined.

An Empirical Study on the Boom Index of STM Journals

According to the indicator system and its weight of the boom index of STM journals, this section empirically studies the boom index by collecting the actual data from CTPCD.

The data sources used in the empirical study include the following:

- III-1 Scientific research and experimental development funds
- III-2 Chinese journal support
- III-3 English-language journal support: data source uses CAS to lead the Quality Science and Technology Journal Financing Project, the Outstanding

Table 20.22 Boom index of STM journals (1999–2013)

Year	1999	2000	2001	2002	2003	2004	2005	2006
CI	11.37	10.36	503.42	550.10	677.09	732.68	847.69	1000.00
Year	2007	2008	2009	2010	2011	2012	2013	
CI	1091.15	1215.36	1304.49	1227.66	1281.44	1270.49	1091.15	

International Journal Award, and the International Influence Journal Financing Project

- III-4 The number of new journals
- III-5 New financial support: data from the National Bureau of Statistics supporting the funding of research and development (1999–2013)
- III-6 Total number of journals
- III-7 Average period print
- III-8 Citation frequency
- III-9 Impact factor
- III-10 Annual index
- III-11 Source literature (quantity)
- III-12 Professional editors: data are from the General Administration of Press and Publication's annual statistical report (2005–2013)
- III-13 Average number of citations
- III-14 Article funding ratio
- III-15 Number of journals in SCI
- III-16 Number of journals in Ei
- III-17 Number of journals in MEDLINE: data sourced from ISTIC and Chinese Science and Technology Statistics and Analysis (2001–2013).

With 2006 as the benchmark, the remaining years of the boom index were multiplied by the weights of the index values of each level and the 2006 values. Thus, the 1999–2013 boom index of STM journals was calculated as shown in Table 20.22.

As can be seen, the boom index of Chinese STM journals showed a rising trend from 1999 to 2013 (Table 20.21). These results should be interpreted with care because there was a lack of data for 1999 and 2000, and because the index score itself is low, the calculation of the boom index is very small and can almost be considered an outlier.

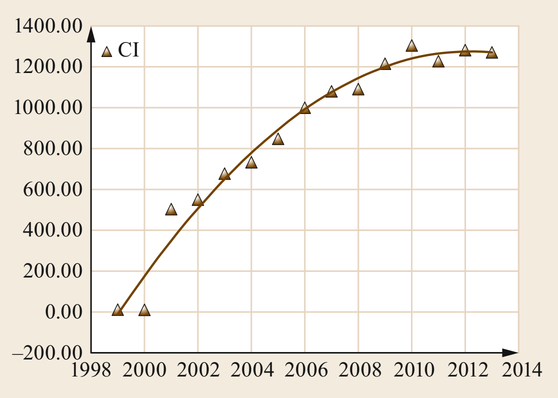


Fig. 20.1 Boom index of Chinese STM journals (1999–2013)

With the 2006 benchmark, the 2006 boom index was calculated at 1000. In the remaining years, the boom index was multiplied by the weights of the indices at all levels and by the 2006 benchmark.

As can be seen from Table 20.22, the 2001 boom index is 503.42, almost half the 2006 value. This result was obtained because prior to 2001, China's STM journals industry boom index was low. Between 2001 and 2006, the boom index maintained rapid growth, with the annual growth rate reaching 26.04%.

Since 2007, the Chinese STM journals industry has stepped up a level, with a boom index of more than 1200 points. Since 2009, the boom index has stabilized at 1200–1300 (Fig. 20.1). However, the pace of growth has slowed accordingly. The average annual growth rate in 2006–2013 was approximately 4.07%. From the above research, the boom index objectively reflects the development of China's STM journal industry.

20.6 The Definition and Application of Comprehensive Performance Scores (CPS) for Chinese Scientific and Technical Journals

STM journals play an irreplaceable role in scientific development and innovation activities, and the scientific progress and sustainable development of journals are the necessary guarantee and key link in the advancement of national science and technology. The research and evaluation results of STM journals are the basic guarantee of scientific decision-making in the management of STM journals and are effective tools to promote the development of STM journals themselves.

Because journal evaluation is a very complicated process that involves a wide range of fields, it is difficult to comprehensively and accurately evaluate the academic level and subject status of a single index. This needs to be evaluated by a comprehensive index, so as to make journal evaluation more objective, comprehensive, and accurate. In the past, the comprehensive evaluation of journals in a particular area was generally divided into two types. One type was experiments based on existing indicators, using mathematical statistical methods or mathematical models for numerical calculation. Methods include principal component analysis [20.36], normalization method [20.37], factor analysis method [20.38], and the rank-sum ratio [20.39]. The advantage of using these methods lies in the objective data that is based on the citation statistic or the literature measure method. The evaluation method is a mature mathematical statistical method. The evaluation process facilitates the public to accept supervision. There are also some limitations in the availability and sustainability of large-scale data. The second approach to the comprehensive evaluation of journals is to adopt AHP or other weighting methods and to summarize several indexes according to the method of normalized weighted calculation. This method has strong operability and applicability, and is often used in large-scale evaluation projects with a large evaluation range, longer time window, complex evaluation indexes, and many evaluation objects. It usually involves the design and creation of a comprehensive evaluation indicator system, using AHP, expert investigation, or other mathematical methods to determine the weights of each index. This is done to find the comprehensive index ranking value and finally to obtain the comprehensive ranking of journal indexes. In the practice of comprehensive evaluation of journals, it is a key technical point to realize interdisciplinary comparison, which is usually achieved by means of standardization [20.40].

20.6.1 Definitions

The comprehensive performance score (CPS) evaluation (hereafter, *comprehensive evaluation score*) of Chinese STM journals is based on the comprehensive evaluation indicator system of such journals. According to the value of periodical indicators in each subject category, the relative position score of each journal in its discipline is calculated, and the comprehensive evaluation score is made up by weighted value of all relative position scores of each indicator.

The annual Citation Report (Core Edition) of China Science and Technology Journals, published by the ISTIC (CSTPCD), regularly publishes measurement indexes of source journals of Chinese STM papers that are cited in other papers [20.41]. Based on that index, the comprehensive evaluation indicator system of Chinese STM journals was developed in 1999 [20.42]. Each index in the evaluation indicator system is not equal to the function of the journal when calculating the total score; thus, different indexes are given different weights. The initial index weight distribution was determined using a Delphi expert investigation method and AHP. Subsequently, with further research and the application of feedback in practice, the new measurement index is continuously supplemented. To assimilate feedback from all walks of life for the evaluation results in a timely manner, the CSTPCD has held more than 20 expert seminars. For those seminars, the institute invited scientific metrology experts, scientists, experts and scholars in science, technology, and journal management to participate. To improve the evaluation system and methods for the current set of scientific and technological journals, the CSTPCD established indicators and assessed weights and adjustments. CPS evaluation is based on the principle of scientific metrology; it comprehensively covers the indexes of evaluated journals. The comprehensive evaluation indicator system of Chinese STM journals is used to classify, divide, and assign different weights to journal indexes; the weighted scores of each index are collected to obtain the total score of the evaluated journals as well as the ranking within the scope of the subject area and within the whole range of journals. Generally speaking, the higher the value of the comprehensive evaluation index, the higher the comprehensive academic quality and influence level of the journal in its subject area.

20.6.2 Index Calculation

In line with the evaluation aims, scope, and angles of different journal evaluations, the index range and corresponding weight distribution are often adjusted when assessing STM journals. Taking the Citation Report of Chinese STM Journals in 2015 (core edition) as an example, the total score of the comprehensive evaluation is the score of the TC, the core impact factor (IF), and the score of six indexes (such as the index (OT), fund thesis ratio (NT), citation rate (RE), and open factor (OP)) according to the relative position of the evaluated journal in its discipline; these indexes are integrated according to certain weight coefficients. The specific algorithm is as follows,

$$\mathrm{CPS} = \sum_{i=1}^{n} \mu_i k_i \,,$$

where n is the index quantity (six indexes) and the weight coefficient of the index, k, is the relative position score of an index of the appraised journal in its discipline. k is calculated as follows,

$$k = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \,,$$

Table 20.23 Calculation of total score of a chemistry journal

Indicator	Value x	Maximum value in chemical science x_{max}	Minimum value in chemical science x_{min}	Relative position score k	Weight μ	Weighted score $^{\mu}k$
TC	2541	4122	100	0.607	26	15.8
IF	1.057	1.954	0.177	0.495	26	12.9
OT	0.87	0.95	0.53	0.810	18	14.6
NT	0.91	0.93	0.24	0.971	10	9.7
RE	51.08	86.45	11.37	0.529	10	5.3
OP	29	60	2	0.466	10	4.7
CPS						62.9

where the value of an index of an evaluated journal is the maximum value of its subject, which is the minimum value of the index of the subject.

The total weight coefficient is designed to be 100; the total score of the comprehensive evaluation is from 0 to 100. The weight distribution of the six indicators adopted in the Citation Report (core edition) of China Science and Technology 2015 was determined using the Delphi expert investigation method. The specific weight distribution was as follows: TC, 26; IF, 26; OT, 18; NT, 10; RE, 10; and OP, 10.

An example appears in Tables 20.23 and 20.24. The index values of a particular chemistry journal in 2014 were compared with the extrema (maximum and minimum) of the corresponding indexes of all chemistry journals. The scores of relative positions of each index were obtained and the total score of the comprehensive evaluation of the journal was weighted.

The same method can be applied for all Chinese science and technology journals to calculate the total comprehensive evaluation score. For 2014, the total comprehensive evaluation score of 38 Chinese science and technology core journals was 62.9, which ranked them fourth in the world. By contrast for 1989, the ranking of all China's scientific and technological core journals was 231.

With respect to calculating the corresponding comprehensive evaluation score, some journals in Citation Report of Chinese STM Journals in 2015 (core edition) covered different disciplines.

Table 20.24 Top ten regions for producing English-language STM journals in China

Rank	Region	Number of English STM periodicals
1	Beijing	167
2	Shanghai	34
3	Hubei	13
4	Jiangsu	12
5	Sichuan	9
6	Zhejiang	8
7	Tianjin	7
7	Jilin	7
7	Guangdong	7
7	Heilongjiang	7

20.6.3 Application

The comprehensive evaluation score has been widely used. In 1999, the CSTPCD established the comprehensive evaluation indicator system of Chinese STM journals. This indicator system has been used to determine the total comprehensive evaluation score of Chinese STM core journals, and is a basis when selecting Chinese STM core journals. For Chinese science and technology core journals (statistical source journals), a dynamic adjustment mechanism—annual evaluation of journals—can be applied by combining quantitative and qualitative methods. Selecting more important or representative subjects can reflect the level of development of various journals. In that selection process, comprehensive journal evaluation is applied to assess new journals. The total score of a candidate journal can be entered into the final procedure by top-third subject ranking (which is exempt from expert inquiry). Alternatively, this approach may be used to eliminate existing core journals, where the elimination is based on the comprehensive score of the journal in its subject area.

The annual Citation Report (Core Edition) of China Science and Technology Journals features a chart, which covers various academic disciplines. The chart presents a comprehensive evaluation of the total number of journals and academic rankings in addition to their total citation frequency, impact factors, and distribution of comprehensive evaluation indicators. In a comprehensive evaluation within an overall ranking table, all journals are sorted according to total comprehensive evaluation score. The values of core influence factors and total citation frequency of each journal are listed. The ranking of all the journals can be used to determine their academic quality and relative position of influence in China. In 2014, the comprehensive evaluation index of science and technology core journals in China was 40.9; the number of comprehensive evaluation indexes amounted to 50 in 517.

In 2005, the National Science and Technology Department began promoting the quality of China's science and technology journals. One step it took was undertaking an overall evaluation and monitoring of those journals so as to stimulate scientific research in dominant disciplines. The department aimed to address problems in existing STM journals, take practical measures to enhance the overall quality and scientific level of STM journals, and promote independent innovation in science and technology in China. It also aimed to encourage the country's top academic journals to attain advanced world levels. The third session of China's top STM journals was published in 2014, and it included 315 publications. The evaluation of those quality journals involved classifying them according to subject area. Journals with higher total comprehensive evaluation scores were assessed more highly.

Since 2009, the CSTPCD has published the results that originally appeared in hundreds of outstanding academic journals in China. Those journals were selected as the most important in their various disciplines. In 2011, the comprehensive evaluation indicator system was used to assess changing trends and the current situation of academic journals in China; the indicator system for journals was modified and the index weights were approved. In 2014, 100 outstanding Chinese academic journals were identified from a total of 315 at the third session of China's top STM journals. With the classification system of Chinese STM core journals, there are 113 categories of disciplines in the natural sciences; thus, it is not possible for every discipline to produce hundreds of outstanding academic journals. During the selection process, the most important academic journals in each subject category are selected as outstanding, taking into account the nature of different disciplines and the size of the journals; an appropriate adjustment to the proportion is then made among the different disciplines.

In recent years, the China Science Association, CAS, Ministry of Health, national press, publications by the SARF, and other technical departments, as well as a number of top industrial associations and local science and technology management departments have made evaluations of academic journals using the comprehensive evaluation indicator system. In addition, journal publication departments and various research projects have made wide use of that indicator system. The total comprehensive evaluation score has emerged as an important indicator, including at the Citic Institute, for the development of scientific and technological journals in China.

In 2012, the CSTPCD and Istic-Elsevier Co., Ltd. established a research center for evaluating journals. One of the aims of that move was to promote China's outstanding STM journals to international data resources systems, such as Scopus and Ei. In selecting recommended journals for international data resources systems to evaluate those Chinese journals, the comprehensive evaluation score is regarded as one of the important objective evaluation bases [20.43].

20.6.4 Discussion

Quantitative indexes established using the citation analysis method can be employed to measure the attributes and distribution of papers published in STM journals. Such indexes can also be utilized to determine the influence and function of periodicals in scientific communication. The main purpose in evaluating and monitoring STM journals in the present study was to assess the evaluation indicator system of Chinese STM journals: it is important that the system accurately reflects the main characteristics of such publications from various angles. The indicator system has to reflect the overall development of STM journals and allow a comprehensive evaluation of each index to be accurately made. Journal editors and publishers may use the indicators to identify their own characteristics and deficiencies, which can assist them in developing the direction of their publications. Based on the quantitative index, the comprehensive evaluation score is an objective standard that reflects the influence level of journals. The score can be used to make an all-inclusive assessment of a journal's overall status with respect to various kinds of STM management.

Using the comprehensive evaluation indicator system of Chinese STM journals developed by the CSTPCD, the present study calculated a number of scientific measurement indexes. This study employed the AHP to determine the weights of important indexes; it evaluated each journal comprehensively and it calculated the total score for each publication. With the comprehensive evaluation indicator system, the total comprehensive evaluation score accommodates differences in the background value of the overall index among different subject areas; thus, it permits an appropriate comparison to be made.

If we assume the IF of physics journal A in 2014 to be 1.066 (average of the maximum (1.954) and minimum (0.177) values of the IF of the subject), the relative position of the journal in the IF index is 0.5. At the same time, the numerical value of the IF of mathematics journal B in 2014 is 0.341 (average of the maximum (0.602) and minimum (0.079) values of the IF of the subject), and the relative position score of the journal in the IF index is also 0.5. There is almost a three-fold difference in the IFs of journals A and B; however, their relative positions in their respective disciplines are the same.

Thus, based on the relative position score and index corresponding to weight, the scores of the two journals' IFs are the same. It is evident that regardless of the value of journal indicators, comprehensive evaluation of a journal permits an evaluation of its relative level within its own discipline; this allows an interdisciplinary comparison among different disciplines.

Some small journals that receive many citations often achieve the highest influence level: almost all the indicators of such journals are greater than those of their peers in the same subject area. The total comprehensive evaluation score signifies the relative level of a journal in its discipline. Accordingly, a single show journal will achieve a higher overall score. In the Citation Report (Core Edition) of China Science and Technology Journals 2015, 15 types of journals are listed, with a total score of over 90. However, with higher-ranked journals, a number of individual indicators (e. g., IFs, total cited frequency) often do not result in a high overall score. This is due to a number of other indicators being low, the degree of reference concentration in the subject being low, or the gap between performance indicators being small. For example, the first journal in a subject area receives a relative position of 1.0. If the second-ranked journal has a number of IFs that are close to those of the first, it will also receive a high score; however, if the difference in the IFs with the second-ranked journal is large, it may receive a relatively low position score despite its ranking.

With overall progress in science and technology and the development of academic publishing and research activities covered by STM journals in China, index selection and weight distribution with the total comprehensive evaluation score will undergo continuous improvement. The score system will be enhanced to meet the management needs of STM journals and various scientific and technological evaluation needs in the country. As part of that process, the CSTPCD will obtain valuable advice and suggestions from the editorial and management sections of scientific and technical journals. The aim of the CSTPCD is to develop a rational, objective operation-evaluation system and tools to provide effective support for the appropriate management of Chinese academic publishing activities and to guide the healthy development of Chinese STM journals.

20.7 Evaluation of English-Language Science and Technology Journals in China

20.7.1 Statistics and Analysis of English-Language Science and Technology Journals in China

The present study was based on a database survey of English-language science and technology (STM) journals in China; the database was created by the ISTIC. This study analyzed the impact indexes of 307 English-language STM journals that were formally published in China. We examined such factors as publication region, year, frequency, and subject area by analyzing the journals' impact in terms of impact indexes, such as impact factor and total number of citations. The number of English-language STM journals published in China has increased and their impact has grown. However, owing to operational resources being diverted elsewhere, English STM journals in China have not undergone large-scale development.

China produces a considerable number of papers in the area of science and technology papers, and the quality of those papers is increasing every year. As of September 2016, Chinese science and technology authors published 1.7429 million articles from 2006 to 2016. In 2015, China ranked second in the world in terms of the output of English-language STM papers; its output represented an increase of 10.2%. English-language STM papers produced in China during the period of 2006–2016 were cited 14.8985 million times—an increase of 15.7%. For that period, China ranked fourth in the world.

China is a major country in terms of producing STM journals; however, it has yet to become a powerful one [20.44]. According to the national press and publication statistics of the State Administration of Radio, Film, and Television (SARFT), China in 2014 published 9966 academic journals, of which 4974 were science and technology publications [20.45]. Compared with 2012, the number of STM journals in 2014 showed an increase of just 0.4%; thus, the number of STM journals in China appears to be reaching stable levels. Scientific and technological cooperation is an important part of research. Together with advances in science and technology and globalization trends, the internationalization of STM journals has become a major development in academic publishing. English-language STM journals have come to assume a very important role as the main platform for the exchange of science and technology information.

The area of China's English-language STM journals has not been extensively researched; when it has been studied, the focus has mainly been on editing, publishing, and distribution. One investigation conducted a statistical analysis of university English-language STM journals used in Chinese universities [20.46]. The study identified 73 journals, but it found that they were not fully representative of current China-produced English-language STM journals. Previous studies have determined that English-language journals published in China lack academic influence as international journals and have low academic indicators [20.47]. The number of English-language STM journals published in various subjects in China is not proportional to the output quantity of international papers covering various disciplines in that country [20.48]. It is imperative for China to publish STM journals in English, and it is necessary to make an accurate, comprehensive analysis of those journals.

In 2001, the ISTIC established a database for English-language STM journals published in China. That database is an important reference tool for management departments and it offers an international retrieval system for accessing the data in such journals. With the rapid development of science and technology and increasing globalization, China's English-language STM journals have likewise quickly evolved; in terms of global literature, the proportion of such journals is growing. As of June 2016, according to statistics of the ISTIC, its database covered 307 domestically published English-language STM journals (with uniform domestic issue numbers). Those journals accounted for 6% of the total number of national natural science and technical journals; that number showed a 1.1% increase over the previous year [20.49]. Compared with STM core journals in Chinese, the publishing frequency of those in English is longer. Among the 307 domestically published English-language STM journals, 123 (42.2%) were published quarterly; they were followed by bimonthly and monthly publications (33.2% and 21.3%, respectively). In 2014, 1989 Chinese-language STM core journals (natural sciences) were published in China; quarterly, bimonthly, and monthly publications accounted for 8.4%, 47.4% and 40.0%, respectively. Compared with Chinese-language STM journals, the number of those in English is low, the operating costs are high, and the publication frequency is low. However, the ISTIC and institutes of higher learning are a major force in establishing domestically published English-language STM journals. In all, 140 such journals are sponsored by the ISTIC, accounting for 45.6% of the total; 201 STM journals in English were supported by two or more organizations (65.5% of the total).

Regional Distribution

This section presents details regarding the regional distribution of English-language STM journals in terms of China's 31 provinces (cities and autonomous regions, excluding Hong Kong and Macao). In 2016, the 307 journals were distributed in 23 regions. Table 20.24 lists the top ten. The greatest concentration of such journals was in Beijing, which accounted for 167 (54.4%), followed by Shanghai, Hubei, and Jiangsu; those four regions accounted for 73.6% of STM journals in English. The number of such journals was related to regional academic exchange activity and also to the number and geographic distribution of scientific research institutions. Only six English-language journals were produced in provincial regions, reflecting the lack of science and technology development in such areas.

Publication History

Before 1949, only three English-language STM journals were published in China. Thereafter, just eight were added in the 30 years up to 1979. After China's economic reform, the number of domestically published STM journals in English developed rapidly. From 1980 to 1990, 79 such journals were established; since then, every decade a large number of those journals have come into being. As of 2011, there were 50 domestically produced English STM journals. The change in the number of journals reflects the progress in China's STM development and also the course of its economic reform.

To promote the international influence and core competence of domestically published STM journals in English, the Chinese Association of Science and Technology launched the International Influence Promotion Program in 2012. The plan called for the investment of almost a billion yuan to establish domestically published English-language STM journals over a three-year period. To further promote the internationalization of China's STM journals, the Chinese Association of Science and Technology, Ministry of Finance, Ministry of Education, SAPPRFT, CAS, and the CAE decided to continue to jointly implement the Chinese Science and Technology International Influence Promotion Program [20.50]. With the second phase of the International Influence Promotion Program (2016–2018), which continues to support new English-language journals in China, it is anticipated that the number of such journals will continue to grow over the next few years.

Interdisciplinary Distribution

With the development of science and technology, academic disciplines are in a state of flux—constantly converging, deriving, and changing. A number of journals cover interdisciplinary fields of scientific research. The database for English-language STM journals published in China and the Citation Report (Core Edition) of China Science and Technology Journals (natural science volume) adopt the same method, which is as follows. According to the main distribution field of each journal, multiple subjects and the interdisciplinary content of journals are classified into two or three subjects. Depending on the discipline, the classification and code (national standard GB/T 13745-2009) and Chinese Book Data Classification (fourth edition) are used for the subject classification; this takes into account the publication history of Chinese STM journals. The source journals are then classified into 113 subject categories [20.51].

According to statistics for 2016, 16 journals were classified into more than one subject category; they accounted for 5.2% of all English-language STM journals produced in China. The 307 such journals were categorized into 96 subject areas (85% of all such categories). The most widely distributed subject area was mathematics, with 19 journals (6.2% of all English-language STM journals); that was followed by physics and engineering, which accounted for 14 and 11 journals, respectively. The three most widely distributed subjects among China's STM journals were engineering and technology, natural science, and medicine.

Statistical Indicators

The source and citation data of English-language STM journals published in China originate from Wanfang Data–Digital Periodicals. Table 20.25 presents the main source indicators for such journals. In that table, *author number of articles* refers to the average number of authors of each paper in the source journal; it is an index of the scientific production capacity of journals. As indicated in that table, the average number of authors for 2014 was 3.98. That figure showed little change from the previous three years. The average number of authors of statistical source journals of Chinese scientific papers (natural sciences) was 4.1 [20.52]. In 2014, the average number of citations of English-language STM journals published in China was 29.9; that is a five-fold increase over 2011. Also in 2014, the average number of citations of Chinese-language STM core periodicals (natural sciences) was 17.1; lower than the figure for English-language journals by 12.8. Compared with statistical source journals, the papers published in China's English-language STM journals have more reference to previous ideas or research results; they are more in line with international standards with regard to bibliographic standardization.

The proportion of papers published by overseas authors in domestic publications compared to the total number of papers in overseas journals is a measure of the degree of internationalization of domestic journals. In 2014, domestically published English-language STM journals showed a slight decline in the number of papers from overseas compared with the previous three years. However, the number of overseas papers showed an increase of more than 50%. In 2012, overseas papers were published in 24 journals, which was less than 50% of the 49 for 2014. In 2014, five overseas papers were published in the following English-language journals: *Frontiers of Biology*, *Current Zoology*, *International Journal of Disaster Risk Science*, *Bone Research* and *Forest Ecosystems*. Attracting more manuscripts from overseas is a problem facing most domestically published journals in English. The proportion of papers that receive funding is an important index of the academic quality of journals. In 2014, the proportion of such papers in English-language STM journals was 65%, which is higher than the average for Chinese-language STM journals (natural sciences) for the same year.

Table 20.26 presents the main citation indicators of English-language STM journals published in China. It is evident that from 2011 to 2014, the total number of citations, impact factors, and the immediacy index

Table 20.25 Main source indicators of English-language STM journals published in China

Indicators	Chinese Journal of English Science and Technology				China Science and Technology core
	2011	2012	2013	2014	CSTPCD 2014 (Natural science volume)
Author number of articles	3.9	3.9	3.8	4.0	4.1
Average number of citations	24.9	27.5	28.4	29.9	17.1
Overseas papers rate	0.23	0.25	0.22	0.16	0.02
Funded papers rate	0.71	0.64	0.44	0.65	0.54

Table 20.26 Main citation indicators of English-language STM journals published in China

Indicators	Chinese Journal of English Science and Technology				China Science and Technology core
	2011	2012	2013	2014	CSTPCD 2014 (Natural science volume)
Total citations	385	380	369	433	1265
Impact factor	0.391	0.370	0.371	0.467	0.560
Immediacy Index	0.092	0.115	0.107	0.132	0.070
Other citation rate	0.84	0.84	0.84	0.85	0.82

showed an increase. This indicates that the impact of such journals is rising. *Total citations* in Table 20.26 refers to the total number of papers of a journal that received citations since the journal's inception: it reflects the extent to which the journal has been used and valued as well as its influence in scientific communication. Owing to the language used in journals and other reasons, the total citation frequency of English-language STM journals is much lower than that of Chinese-language ones. In 2014, English-language journals were cited 433 times; that means a 64 times increase comparing with 369 times in 2013.

Impact factor (IF) refers to the number of citations that appeared in the two years prior to a journal's evaluation. It is a measure of the journal's academic impact. The change in the IF for 2011–2014 for English-language STM journals and the Chinese-language STM journals is indexed by CSTPCD. Noticeable is that the IF of Chinese-language STM papers displayed a gradual increase during that period, while the IF of English-language journals was relatively steady. In 2014, a major improvement occurred: the IF of English-language journals in that year was 0.467, which was 25.9% higher than the 0.371 in 2013.

The immediacy index refers to papers published the same year as a cited paper: it characterizes the rate of immediate response indicators. The annual index of English-language STM periodicals published in China in 2014 was 0.132. It is evident in Table 20.26 that the instant reaction rate of such journals was slightly higher than that of the *Journal of Statistical Sources* (natural sciences) among Chinese-language STM journals. The quoted rate refers to the total citation frequency divided by the number of other citations accounted for in the ratio. The Chinese English-language STM periodicals are relatively stable in this rate, maintained at about 0.85, slightly higher than China's scientific and Technological paper Statistical Source Journal (natural science part) of his cited rate.

International Impact

International cumulative citation data for English-language STM journals published in China were derived from the Web of Science (SCI) database of the Thomson Reuters Group for 1995–2016 (as of June 2016). As of June 2016, such journals published in 2013–2014 were cited more than five times in 19 different journals. In 2016, only seven of the indicators exceeded five times. The number of English-language journals has increased greatly since June 2016; this indicates that the international influence of such journals is rising. The number of journals cited more than 300 times in 2016 was 13; that figure was seven in 2014.

According to the 2015 Journal Citation Report (retrieved in June 2016), among the 307 English-language STM journals published in China, 133 (43.3%) were included in SCI. Nine of China's English-language journals appeared in the top 25% of SCI journals in 2015; that compares with six in 2012; among those, the *Asian Journal of Andrology* (medicine specific to men) ranked first. Among China's STM journals, 133 covered 86 subject areas; among those, three were classified into four categories, ten into three categories, and 37 into two categories.

Conclusion

With improvements in China's academic quality and international influence, domestically produced English-language STM journals have undergone rapid development. The number of such journals has achieved basic stability: it has shown an increase of 1.1% compared with the previous statistical year. As of June 2016, 307 English-language STM journals were being produced in China (with a uniform domestic issue number). Such journals are an important platform for demonstrating the level of China's science and technology development and also enhance international cooperation.

Producing English-language STM journals demands high resource costs. The production of half of such journals in China is concentrated in Beijing; the next highest areas of production are in Shanghai, Jiangsu, and Hubei. Some scientific papers originate in China's provincial regions; however, some economically developed regions do not produce an appropriate number of English-language STM journals.

The key to an English-language journal's success is that it should be an internationalized operation that is able to attract outstanding papers from overseas. In 2014, overseas contributions accounted for more than 50% of papers in the case of 49 journals published in China. Compared with the previous statistical year, that is an increase of 25 journals. However, the overall level did not improve, and there is considerable room for development.

Among the 307 English-language STM journals produced in China, 85% cover 96 subject areas. The most common subject area is mathematics. English-language journals need to develop a clear orientation. They need to expand their international audience, strengthen their international operations, shorten their publication cycles, and improve their quality. Currently, the publishing cycle of China's English-language STM journals is relatively long: 42.2% of such journals are quarterlies.

The total citation frequency and influence factor of English-language journals have improved compared with the previous statistical period. The international cumulative index of the journals has also improved. As of 2016, 133 English-language STM journals were included in SCI: nine featured among the SCI's top 25% of journals, which was an increase of three over the previous statistical period. That is an indication of the growing influence of English-language STM journals produced in China.

20.7.2 Communication Value of China-published English-Language Academic Journals According to Citation Analysis

Introduction

The Bible has been translated into more than 1800 languages, but there is no definitive figure as to the total number of languages in the world. People use thousands of different languages. The communication barriers posed by language differences clearly affect cultural exchange and information sharing among different countries and nationalities. For barrier-free exchange among nations, there would seem to be only two ways to design and create a world language for all to learn based on natural human language. In this regard, linguistics has created a specialized discipline, termed Esperanto linguistics (Interlingua). Many studies and attempts have been made towards that end, but it would appear to be a dead end. The second option is to make a natural language to fulfill the role of Esperanto and become the mother tongue or second language of all people [20.53]. In view of the current world situation with respect to language development, English is widely learned and clearly used as Esperanto was created to be used. Notably, in activities related to international STM research and exchange, English plays a central role.

Abram de Swaan describes the relationships of various languages in the world in terms of supercentral, central, and peripheral languages [20.54]. A supercentral language is analogous to a star surrounded by planets (central languages), and the planets have their own moons (peripheral languages). When native speakers of one language learn another language, they often choose a broader language within the system of languages. Each supercentral language is associated with a number of users of central languages. With time, a large number of central language groups become linked to an oversized language group among the supercentral languages as a result of multilingual users.

The importance and position of a language in the system of world languages can be quantitatively assessed by calculating the communicative Q value of a language by means of the popularity index and central index. One study found that 12 languages—Arabic, Chinese, English, French, German, Hindi, Japanese, Malay, Portuguese, Russian, Spanish, and Swahili—constituted the supercentral languages. With the possible exception of Swahili, the number of speakers of supercentral languages is over 100 million each.

Among the supercentral languages, English is at the very center of the entire linguistic galaxy. Accordingly, *de Swaan* refers to English as the world's only hypercentral language [20.54]. Globally today, English is the most common language in science and technology. Most STM journals—especially high-quality journals—are published in English [20.55]. With the broadening of the international vision of Chinese researchers and rapid development in linguistics research [20.56], a clearer, deeper understanding has emerged in China of the role of English in science and technology. From the growing number of English-language STM journals produced in China, it is evident that English is playing an increasingly important role in its publishing and exchange activities in STM.

According to the English-language STM journal database, developed by the ISTIC, 212 English-language STM journals were published in China in 2010. The number of such journals has increased annually; the subject areas covered by those journals has also grown. Table 20.27 presents the historical development of China's English-language STM journals. It is evident that the number of such journals was very small before 1980. Over the past 30 years, the number has surged and that growth trend has been maintained [20.57].

As a result of the increasing volume of English-language STM journals produced in China, overseas scholars have devoted greater attention to the development and study of such publications. Those authors have analyzed the role of those journals with respect to research and the exchange of information related to STM in China. However, the conclusions and opinions of those scholars are not wholly consistent. Some authors believe that only by producing quality English-language STM journals can China become truly internationalized in an academic sense. For example, some

Table 20.27 Historical trend of new English-language STM journals published in China

Period	Before 1949	1950–1960	1961–1970	1971–1980	1981–1990	1991–2000	2001–2009
Number of new Chinese and English STM periodicals	2	10	2	6	75	62	54

authorities view language as the greatest obstacle: they regard it as an impediment to the internationalization of Chinese STM and that in "the internationalization of STM periodicals, English periodicals should be first" [20.58].

Some scholars believe that the number of English-language STM journals published in China falls below the country's requirements. Those authors consider that China's journals need to compete with other STM publications produced around the world; they believe that China has the potential for vigorous development and that it has a great role to play [20.59].

A number of studies have suggested that although in China the circulation of English-language STM journals is often lower than that of their Chinese-language counterparts, the impact on the readership is quite similar [20.60]. One report identified 11 characteristics of a world-standard journal, the first of which was publication in English; the second was a fully internationalized peer-review system [20.61]. However, other studies have found that there appears to be no link between international journals that are successful in China and use of the English language. One investigation conducted a comparison of various indexes of Chinese- and English-language journals in the field of materials science using an international retrieval system. The study determined that such features as the impact factor of English-language journals did not offer any advantages; it identified weaknesses as small publication scale and narrow area of influence [20.62]. Some scholars believe that publishing in English is not necessary for an international journal. They consider that the use of English or Chinese in STM journals published in China constitutes a strategic choice; whether the use of English is expected to play a key role depends on whether the journal conforms to its anticipated development strategy [20.63].

The basis for China's STM journals and the main object of publication are the vast number of Chinese people engaged in STM; such workers blindly pursue the study of English and forget the original purpose of those journals [20.64]. Some scholars believe that Chinese-language journals have a strong regional characteristic: they come mainly from China. Thus, language alone does not determine the influence of Chinese STM journals or account for their weak international competitiveness [20.65].

From the above, it is evident that current research on the role of English-language STM journals produced in China has been based on qualitative analysis methods. No objective, unified quantitative evaluation system has been applied to the topic, and so different authors have arrived at different conclusions. To attain a more accurate judgment on the value of such journals and provide objective data support with regard to their future development, the present study applies the ideas of de Swaan. This study examines the Q value of language communication, and it defines a new evaluation index using the Q value of English-language STM journals published in China.

Calculation Method of Exchange Value of English-Language STM Journals

This study takes into account differences in the prevalence of English among different subject areas. It does so with respect to the level of academic influence of journals and their degree of internationalization. This investigation determines the exchange value of English-language STM journals using quantitative research methods; it demonstrates the function and position of such journals in their subject areas using those methods.

In determining the exchange value of these journals, this study makes use of two indexes related to popularity and centrality as defined by de Swaan with regard to the language communicative value index Q [20.66]. This investigation also takes into account factors related to the international influence of such journals. Studies have shown that the number and proportion of papers cited in Chinese journals are generally lower than in overseas journals; accordingly, the international academic influence of STM journals published in China is particularly important.

The exchange value of English-language STM journals produced in China consists of three parts: popularity of English in the subject area; center of academic influence; and degree of international diffusion. For journal j in subject area i, the exchange value indicator is defined as follows,

$$Q_{ij} = \sqrt{P_i^2 + C_{ij}^2 + D_j^2}\,, \tag{20.1}$$

where Q_{ij} signifies the exchange value of the journal, P_i the popularity of English in the journal's subject area, C_{ij} the academic influence center of the journal in its discipline, and D_j the degree of international diffusion with respect to the journal's academic influence.

With the aim of an index calculation with good feasibility and validity, the indicators adopted in the present study are defined and calculated using citation analysis in scientific metrology.

English Popularity of Subjects. A notable feature of language is that the more people who use it, the greater its exchange value [20.67]. As a result, we may conclude that the greater the popularity of English in a scientific research area or the wider the use of English as an academic communication tool, the greater the importance of English-language STM journals.

The prevalence of English in a discipline can be examined by measuring the extent to which papers in that field use English. In terms of research results, the higher the absolute number of papers in English and their relative proportion, the greater the need for the communicative language to be English—and the greater the role and exchange value of English journals in that field. For a subject i, the equation for the English prevalence index is as follows,

$$P_i = \frac{N_i(\text{Engl. Ref.})}{N_i(\text{Total Ref.})} , \quad (20.2)$$

where P_i denotes the English prevalence of subject i, N_i(Engl. Ref.) the number of references in papers published in the particular journal on subject i, and N_i(Total Ref.) the number of all references in all journals to subject i.

Academic Impact Center of a Journal in its Discipline. Irrespective of whether a journal is in Chinese or English, the more it is used in academic exchange, the more it will be regarded as authoritative and will assume a guiding position; this will be reflected in the number of citations in other journals. Thus, the academic influence of a journal signifies its authority and academic position among other journals in a particular subject area.

In a certain field, mutual citations among journals can be used to represent the relative position of each; mutual citations reflect the academic exchange relationship and status among journals [20.68]. To determine the authority and position of a journal, the present study calculated the number of citations of a journal by other journals in the same subject area as well as the proportion of the maximum number of citations in that subject area. For subject area i, the academic impact center of journal j is defined as follows,

$$C_{ij} = \frac{N_{ij}(\text{Citation from others})}{\max_{j=1\approx n} N_{ij}(\text{Citation from others})} , \quad (20.3)$$

where C_{ij} signifies the academic impact center of journal j within subject area i, n the number of journals in subject area i, and N_{ij}(Citation from others) the number of citations of journal j by other $n-1$ journals in subject area i.

With this index, when the academic influence of the journal is determined, the number of citations is not taken into account. A reasonable amount of self-citation by a journal is normal in academia; it also reflects a journal's influence. However, in the present study, the emphasis is on measuring a journal's exchange value: the journal's external influence. In addition, self-citation by a journal is often susceptible to human factors and it does not necessarily reflect the influence of a paper in an objective manner. As a result, self-citation is not included in the present statistical analysis.

International Diffusion Degree of Academic Influence of Journals. The present study conducted an analysis of the exchange value of English-language STM journals published in China. The analysis included the exchange value in domestic disciplines. This study also examined the international exchange value of those journals: it investigated the achievements of Chinese STM in international academic circles and the promotion of international academic exchanges. Therefore, in designing the index for exchange value Q, this study added a third part: the international diffusion degree of academic influence of periodicals, which is reflected in the number of citations in international journals and number of citations in domestic journals. The equation for determining the international diffusion degree of academic influence of a journal j is as follows,

$$D_j = \frac{N_j(\text{NIC})}{N_j\,(\text{NIC}) + N_j(\text{NDC})} , \quad (20.4)$$

where D_j signifies the international diffusion degree of academic influence of journal j, N_j (NIC) the number of citations of journal j in international retrieval systems (i. e., the number of citations in international journals), and N_j (NDC) the number of citations of journal j in domestic retrieval systems (i. e., the number of citations in domestic journals).

Calculation of Exchange Value of 45 English-Language STM Journals

The present study examined 45 English-language STM journals included in JCR 2010 and the Citation Report (Core Edition) of China Science and Technology Journals 2011. This study used the CSTPCD of the ISTIC and the Scientific Citation Index (SCI) database of Thomson Reuters in the United States, respectively for domestic and international citation statistics in calculating the exchange value Q of the sample periodicals [20.69, 70]. The classification of subjects was based on the classification of the Citation Report (Core Edition) of China Science and Technology Journals 2011. All the citation analyses and calculations were based on statistics for 2010. The results of the calculation appear in Table 20.28.

As shown in Table 20.28, the Q value of the 45 journals was 1.475; the minimum was 0.676, median 1.027,

Table 20.28 Exchange value Q for 45 English-language STM journals published in China in 2010

Title j	Field i	P_i	C_{ij}	D_j	Q_{ij}
Acta Biochimica et Biophysica Sinica	Biology	0.802	0.046	0.888	1.197
Acta Mathematica Scientia	Mathematics	0.837	0.853	0.464	1.282
Acta Mathematica Sinica English Series	Mathematics	0.837	0.717	0.732	1.323
Acta Mathematicae Applicatae Sinica	Mathematics	0.837	0.085	0.775	1.144
Acta Mechanica Sinica	Mechanical	0.597	0.272	0.625	0.906
Acta Metallurgica Sinica	Metallurgical engineering technology	0.645	0.095	0.879	1.094
Acta Pharmacologica Sinica	Pharmacy	0.574	0.203	0.731	0.951
Advances in Atmospheric Sciences	Atmospheric science	0.468	0.311	0.557	0.791
Asian Journal of Andrology	Clinical	0.578	0.111	0.736	0.942
Cell Research	Biology	0.802	0.074	0.842	1.165
Chemical Research in Chinese Univ.	Chemical	0.914	0.060	0.660	1.129
China Ocean Engineering	Marine science	0.599	0.067	0.457	0.756
Chinese Annals of Math. Series B	Mathematics	0.837	0.456	0.658	1.158
Chinese Chemical Letters	Chemical	0.914	0.156	0.747	1.191
Chinese Journal of Aeronautics	Aerospace science and technology	0.567	0.198	0.372	0.706
Chinese Journal of Cancer Research	Oncology	0.802	0.020	0.512	0.952
Chinese J. Chemical Engineering	Chemical engineering	0.583	0.181	0.617	0.868
Chinese Journal of Chemical Physics	Physics	0.898	0.014	0.552	1.054
Chinese J. Oceanology and Limnology	Marine science	0.599	0.116	0.618	0.868
Chinese Journal of Polymer Science	Chemical	0.914	0.037	0.721	1.165
Chinese J. Structural Chemistry	Chemical	0.914	0.107	0.626	1.113
Chinese Medical Journal	Medical synthesis	0.520	0.346	0.560	0.839
Chinese Optics Letters	Physics	0.898	0.093	0.490	1.027
Chinese Physics B	Physics	0.898	0.594	0.381	1.142
Chinese Physics C	Physics	0.898	0.047	0.390	0.980
Chinese Physics Letters	Physics	0.898	0.330	0.642	1.152
Communications In Theoretical Phys.	Physics	0.898	0.132	0.633	1.107
Insect Science	Biology	0.802	0.034	0.683	1.054
J. Computer Sci. and Tech.	Computer science and technology	0.665	0.056	0.227	0.705
Journal of Environmental Sciences	Environmental science and technology	0.498	0.180	0.674	0.857
Journal of Genetics and Genomics	Biology	0.802	0.166	0.191	0.841
Journal of Geographical sciences	Geography science	0.375	0.002	0.890	0.966
Journal of Integrative Plant Biology	Biology	0.802	0.353	0.306	0.928
J. Mater. Sci. & Tech.	Materials science	0.752	0.153	0.743	1.068
Journal of Molecular Cell Biology	Biology	0.802	0.030	0.286	0.852
Journal of Natural Gas Chemistry	Energy science and technology	0.350	0.005	0.732	0.811
Journal of Rare Earths	Materials science	0.752	0.302	0.650	1.039
J. Wuhan Univ. Tech. Mater. Sci. Edition	Materials science	0.752	0.068	0.746	1.061
J. Zhejiang Univ. Science A	Journal of universities	0.577	0.057	0.679	0.893
Molecular Plant	Biology	0.802	0.184	0.359	0.898
Particuology	Chemical engineering	0.583	0.048	0.618	0.851
Pedosphere	Agronomy	0.361	0.100	0.563	0.676
Res. in Astronomy and Astrophysics	Astronomical	0.988	0.294	0.618	1.202
Trans. Nonferrous Metals Soc. China	Metallurgical engineering technology	0.645	0.476	0.601	1.002
World Journal of Gastroenterology	Internal science	0.826	1.000	0.702	1.475

and average value 1.004. From the distribution of the Q value, it is evident that there were only four journals with a Q value of more than 1.2; 19 journals with a Q value between 1 and 1.2, accounting for 42% in 45 periodicals; 17 journals with a Q value between 0.8 and 1, accounting for 38% in 45 periodicals; and 5 journals with a Q value of no more than 0.8. It is clear from the distribution of the Q value of the 45 journals (Fig. 20.2) that the calculated results of the sample show approximately a normal distribution.

In terms of disciplines, the 45 journals cover basic research in life sciences and engineering technology. However, there is no obvious difference in the evaluation results among the various disciplines.

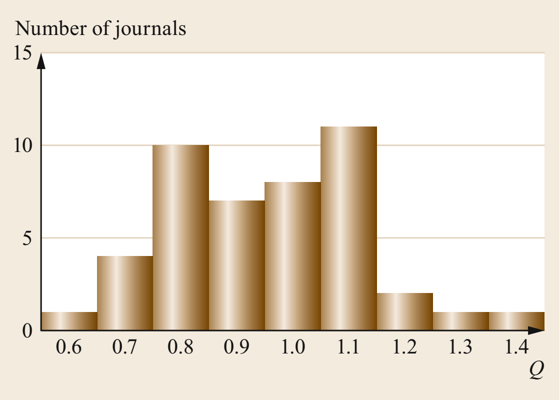


Fig. 20.2 Distribution of Q value of 45 English-language STM journals

It appears that journals in gastroenterology and cell research as well as other publications with distinctive characteristics and higher academic influence showed a good performance in terms of exchange value and Q value. At the same time, it is evident that the Q values of such journals as *Pedosphere* and *Chinese Journal of Aeronautics*, which have a clear regional bias and represent closed disciplines, are relatively low [20.71]. Therefore, it can be considered that the exchange value index Q of English-language STM journals published in China has a certain degree of accuracy; it may be used to characterize the actual role of such publications with respect to China's scientific research and publication exchange activities.

Conclusion and Discussion

Studies in linguistics have shown that English has become the only language in the world with the status of a hypercentral language. It is necessary to fully understand and use this communication tool in STM. The exchange value Q of English-language STM journals produced in China is a measure of the role of such publications in Chinese science and technology with respect to academic research and exchange activities. The exchange value Q can be used as an objective, practical quantitative analysis tool to track the development of such journals.

As evident in (20.1), the exchange value of a journal is expressed by the three coordinates of popularity (P), center (C), and international diffusion degree (D); the length of the vector formed by 0 points to the position of the journal. The relative differences in the results of those three coordinates can be likened to different angles produced by a journal in a three-dimensional coordinate space.

In subject areas where there is higher prevalence of English, Chinese STM journals are more likely to use that language for ease of communication. The prevalence of English (p) can be regarded as a measure of the international strength of a subject area. The prevalence of English can also be regarded as reflecting the demand for English-language journals as well as a reflection of the potential exchange value of such journals produced in China. The higher the proportion of Chinese researchers who access English-language literature, the greater the exchange value (potential exchange value) of English-language journals in China.

In designing the exchange value index, the difference between disciplines is considered a component. Thus, the data obtained using that index are not affected by differences in subject area. The exchange value Q of English-language STM journals can therefore be used for comparisons across disciplines.

The design of the exchange value index is based on the situation of English-language STM journals produced in China. It reflects the role played by Chinese scientific research activities. The domestic journal citation database of CSTPCD and international journal citation databases, such as SCI, are used to calculate Q. For other countries, especially non-English-speaking countries, it is impossible to use that index to calculate the exchange value of journals if they lack a complete domestic citation database like China's. Therefore, the exchange value index Q cannot be used in international comparisons. However, in a follow-up study, it will be possible to consider the value of Chinese-language STM journals and compare them with English-language journals produced in China.

The exchange value index reflects and evaluates the function of English-language STM journals produced in China from one aspect. The index cannot completely replace conventional measurement indexes of scientific journals, such as influence factor and total cited frequency. However, the exchange value index can be used as an extension to supplement the current comprehensive academic indicator system of English-language STM journals produced in China [20.47].

20.7.3 Atomic Structure Model for Evaluating English-Language Scientific Journals Published in Non-English Countries

Introduction

English-language scientific journals published in non-English-speaking countries account for only a small fraction of all scientific journals published in those countries. However, they play a particularly important role in connecting scientists whose native language is not English with their peers around the world and in improving international collaboration through scientific research. In China, for example, although over 5000 scientific journals are published (ranking it second after the United States in terms of number of journal titles published), its English-language journals amount to only about 4% of the whole [20.72]. As is well known, English is the most widely used language in almost all fields of science, technology, and current international economic and social affairs; there is a tendency to accept English as an official international language. Most scientific journals, especially the most outstanding ones, are published in English [20.73]. It is axiomatic therefore that international scientific communications and publications follow the existing trend. Accordingly, scientific journals published in English have become the major channel for international exchange and cooperation. Over 100 000 papers by Chinese scientists are indexed annually in Science Citation Index Expanded (SCIE); 14% of those papers appeared in English-language scientific journals published in China [20.74].

China aims to become a substantive member of the world scientific community. Thus, it has become important to make those few Chinese journals published in English better and more attractive to international scientists as places for publishing their academic findings. The same challenge is probably faced in other countries where English is not an official language.

In general, evaluating something is a good step towards improving it: the grades or marks accorded as part of a logical evaluation process and related to systematic indicators denote the advantages and disadvantages as well as the strengths and weaknesses of the matter under evaluation in a significant, precise, all-encompassing manner. Thus, evaluation is akin to the role of diagnosis, upon which prescription and amelioration are founded.

English-language scientific journals published in non-English-speaking countries have certain characteristics that make them different from most non-English journals. Those characteristics result in a lack of an accurate, suitable evaluation model and indicator system. Such English-language journals focus on international readers and authors; that is why their sponsors and editors made the decision to operate them in a country where English is not an official language. In China, it has been observed that international journals usually acquire more citations than domestic ones [20.75]. It is understandable why English-language journals in China do not attract many domestic authors and readers, who pay more attention to journals and articles in Chinese owing to the greater convenience of obtaining information and producing articles in their native language. Some scientific evaluation models or systems have been established for journals published in China, and English-language journals consistently obtain lower scores than most Chinese journals [20.76]. Most English-language journals deal with more peripheral fields, according to the mutual citation network, which groups journals by subject areas [20.77]; however, such evaluation results may not reflect the real academic level and impact of such English-language journals.

Most English-language scientific journals published in China are very new and limited in number and circulation; thus, they do not yet have sufficient strength to exert a wide influence and strong impact on the world scientific community. Established in 1887, China's first English-language scientific journal was entitled *China Medical Missionary Journal*. However, a recent survey conducted by W. Xu revealed that very few other English-language journals appeared in the almost 100 years that followed until the 1980s [20.78].

According to a comparison of the number of new English-language scientific journals published in China per decade, we can find that that the great majority have appeared in the last three decades; more than a quarter emerged after 2000. Those journals are so new that it would be inappropriate to evaluate them using the same standards that can be applied to other older Chinese-language journals.

More than half of China's English-language journals are not covered by SCIE. This means that most of them do not have an ISI impact factor or other indicators calculated by ISI JCR, the most popular international index of scientific journals [20.79]. Hence, neither the existing domestic evaluation system nor the normal international evaluation system is entirely appropriate for China's English-language journals.

It is possible that some of those journals may fall into a vicious cycle in the future if the current evaluation models continue to be adopted. That is because less influence leads to a lower evaluation score, a lower score leads to fewer submissions, and fewer submissions lead to less academic value and poorer impact of the journal. To address this problem, a new model, which resembles the structure of an atom, has been designed as described below to evaluate China's English-language scientific journals.

Atomic Structure Model

Characteristics of the Model. The most important achievement of John Dalton, a British chemist and physicist in the 1800s, was the atomic theory. Some scientists, including J.J. Thomson, Jean Baptiste Perrin, Hantaro Nagaoka, Niels Bohr, and Ernest Rutherford, conducted extremely fruitful research over a number of years in developing that theory and producing the universally acknowledged atomic model. As devised by Bohr and Rutherford, the atomic model is a structure

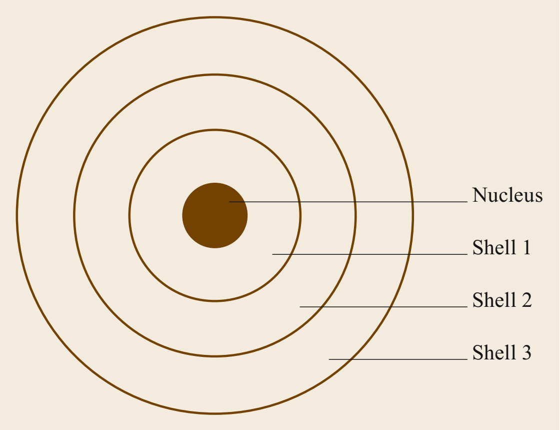


Fig. 20.3 Atomic structure model

that consists of various shells around the nucleus. In the present study, the atomic structure model (ASM) can be represented by the image shown in Fig. 20.3. There, three shells surround the nucleus, and the relationship (degree of interaction) between the shells and the nucleus gradually recedes with distance. The five characteristics of the ASM are as follows:

(1) There is only one nucleus in the system
(2) Three shells are located around the nucleus
(3) Each shell has a spherical surface
(4) Both shell–nucleus and shell–shell interactions occur
(5) The electrons on the three shells orbit around the nucleus.

Use of the ASM in Communication. Yongtao used the ASM in his study of an online communication model. According to his idea, the core of the model is human subjects, who actively push information [20.80]. All objects of information to be communicated by the subjects are distributed on the three shell layers; each shell corresponds to a different effect of communication or the environment of websites and human behavior. To maintain the balance of the whole system there are at the same time complex interactions among all elements in this model.

Scientific journals can be regarded as a branch of communication. Therefore, the ASM would appear to be useful in evaluating scientific journals.

Use of the ASM in Evaluating China's English-Language Scientific Journals. China's English-language scientific journals have some characteristics that match those of the ASM. The essential element determining a journal's academic quality is its core value. The core value is signified by the core (C). Around the core value, there are three classes of individuals located in order of decreasing relationship with the journal. They are somewhat similar to the electrons orbiting in three shells around the nucleus. As Fig. 20.4 shows, the three groups are the editors, authors, and readers. They appear, respectively, as Shell 1 (S1), Shell 2 (S2), and Shell 3 (S3) in that figure.

Editors, including chief editors, associate editors, and members of the editorial board, are those closest to the journal's core; thus, they are located in S1. In their work, they follow principles that are central to the production of a good journal, such as evaluating the quality of submissions and ensuring that the English is of an appropriate level. China's English-language journals are similar to those published overseas and are generally regarded as different from other journals produced in that

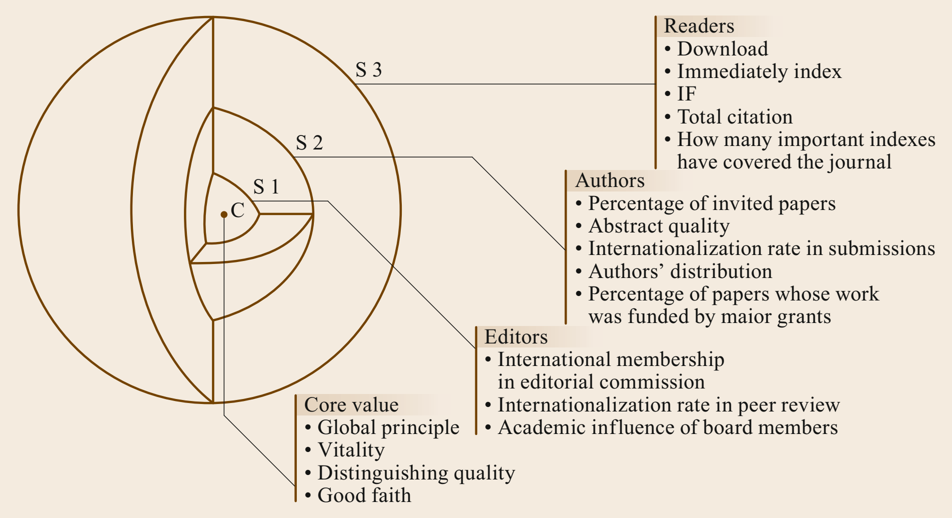


Fig. 20.4 Atomic structure model and indicators for evaluating China's English-language scientific journals

country. Therefore, to operate such English-language journals, it is a prerequisite that the editors have a global view and the capability for international exchange and collaboration. This point has to be examined in the evaluation.

S2 consists of authors whose articles are published in a journal. Clearly, a scientific journal's content mostly consists of academic articles from a large number of authors. Thus, the contact of each author with the journal may not be very frequent if they contribute only one or two submissions a year. The ability to claim a greater number of international authors is a positive attribute for international journals [20.81]. Therefore, it is necessary that the evaluation indicators reflect the quality of both authors and articles.

The most distant shell, S3, is that of the readers. Readers normally have a loose relationship with the journal's core, and direct communication between the two is ordinarily a one-way process. The evaluation of a given journal by readers offers an indirect way of observing the journal's quality by identifying and assessing its academic effect. Consequently, the following questions relating to citation analysis have to be answered to evaluate the response from readers to the journal in S3: how many readers have downloaded papers published in the journal, and what are the journal's citation indicators, such as impact factor and total number of citations?

Interaction between the shells and the nucleus occurs, just as it does among the various shells. A journal with a strong core value usually invests more resources in the journal, including human resources and financial support. In that way, the editorial team can be more effective in soliciting higher-quality submissions from good authors, and this in turn attracts more readers and more citations. This situation is similar to that of an electron's transition between shells in the ASM in that the status of all those involved with the journal changes according to the shell: a reader may become an author, and an author may become active as an editor of the journal. With a quality journal undergoing positive growth, there will be a good deal of transition from S3 to S2 and from S2 to S1: more readers will want to submit manuscripts and more editors will be required to deal with the increasing flow of manuscripts. Conversely, if a journal is not doing so well, some of its former authors will stop submitting new manuscripts, and the number of readers will also start to dwindle.

Based on the ASM for evaluating English-language scientific journals published in China, a series of tentative indicators was developed after dozens of information scientists and academic journal editors brainstormed the issue. Some of these tentative indicators were discarded owing to limited data availability or overlap with other indicators. Finally, 17 indicators were established for this evaluation model (Table 20.29).

Results

To test the usefulness of the ASM for evaluating China's English-language scientific journals, we selected 18 such journals as our sample (Table 20.30). First, we calculated the quantitative indicators for those journals, which are indicated by asterisks in Table 20.29; for the evaluation, we obtained print copies of the journals. Then, for each journal, a group of reviewers (consisting of three to five experts, including scientists in the journal's academic field and senior journal editors) were invited to make a conclusion according to the ASM indicator system. Each journal's core and three shells were rated as A (excellent), B (good), C (fair), or D (poor) through a combination of this quantitative and qualitative analysis. For example, *Journal of Computational Mathematics* was rated as A in S1, which means that editorially this journal is strong according to the three indicators listed in Table 20.29. Although the journal boasts an excellent performance, it scores only C in S2 and S3; this means that the journal's editors

Table 20.29 ASM indicators for evaluating English-language scientific journals published in China

ASM	Core value	Editors shell 1	Authors shell 2	Readers shell 3
Indicators	Global principle	International membership in editorial commission[a]	Percentage of invited papers[a]	Download times according to databases online[a]
	Vitality	Internationalization rate in peer review[a]	Abstract quality	Impact factor[a]
	Distinguishing quality	Academic influence of board members[a]	Internationalization rate in submissions[a]	Immediacy index[a]
	Good faith		Authors' distribution[a]	Total citations[a]
			Percentage of papers whose work was funded by major grants[a]	Number of important international indexes that have covered the journal[a]

[a] Quantitative indicator

Table 20.30 Evaluation of 18 English-language scientific journals published in China

Title	Core	S1 shell	S2 shell	S3 shell
Advances in Atmospheric Sciences	B	B	C	C
Biomedical and Environmental Sciences	B	C	D	D
Chinese Medical Journal	B	D	D	B
Chinese Physics C	B	D	D	C
Journal of Computational Mathematics	B	A	C	C
Journal of Environmental Sciences-China	B	C	C	C
Journal of Genetics And Genomics	B	C	C	C
Journal of Integrative Plant Biology	B	C	B	D
Journal of Iron And Steel Research Int.	C	D	C	D
Journal of Univ. of Sci. and Tech. Beijing	C	C	D	B
Science in China: Mathematics	C	D	B	C
Science in China: Chemistry	C	C	C	C
Science in China: Life Sciences	C	C	D	C
Science in China: Earth Sciences	C	D	C	C
Science in China: Technological Sciences	C	D	B	C
Science in China: Information Sciences	C	D	D	C
Science in China: Physics Mech. & Astronomy	B	C	D	C
World Journal of Gastroenterology	B	C	C	B

are unable to attract top authors and expand their readership. The publisher or editor-in-chief of this journal should perhaps consider this situation and take appropriate steps. Another example is the journal *Chinese Physics C*, which scored D in S1 and S2; this means that the combined power of its editors and contributing authors is very weak in terms of the indicators defined above. Therefore, from this evaluation result, it would appear that the journal needs to make great efforts to bolster the capacity of its editors so as to attract better authors and improve the quality of its submissions.

Conclusion

The primary innovation in the present research was the use of the ASM to evaluate China's English-language scientific journals. Both indicators and layers of indicators in addition to the relationship and transitions between indicators and layers were considered by means of the ASM, a concept adapted from the field of communications.

Another original feature of the present study was considering the interaction among the shells in the ASM: these signify the individuals working for or related to a scientific journal. The changes in a journal's editors, authors, and readers could reflect its stage of development and possible future trends. However, the lack of data from the journals themselves meant that there was no possibility of evaluating such details in the present study.

Unlike the situation in some countries around the world, all Chinese journals are managed by public academic organizations, such as universities, research institutes, and academic societies. As the sponsor—more often than not the only sponsor—of scientific journals, those organizations have the competency and absolute responsibility to formulate a vision for a journal's development and carry it out. This aspect of the sponsor, or perhaps owner, is an important element with regard to the quality of a journal published in China; however, it was not considered in the present study.

With necessary adaptations, the indicators designed for China's English-language scientific journals could also be used to evaluate English-language journals published in other countries where English is not an official language. For example, English-language scientific journals published in Japan face many similar problems to those produced in China [20.82].

The ASM evaluation idea could also be applied in evaluating all scientific journals irrespective of language or country of publication: all journals share the same essential elements of core value, editors, authors, and readers. Further study is required to optimize the indicator system used with the ASM and verify it in a large sample.

References

20.1 J. Qiu: *Bibliometrics* (Science and Technology Documentation, Beijing 1988)

20.2 R. Qian: *Core Periodicals and Evaluation of Periodicals* (China Communication Communication, Beijing 2006)

20.3 J. Qiu, E. Li: The theory, practice and development trend of Chinese periodical evaluation, Digit. Libr. Forum **3**, 1–12 (2007)
20.4 M. Lai, P. Qu, K. Zhao: The origin and core elements of journal assessment, J. Chongqing Univ. Soc. Sci. Ed. **3**, 67 (2009)
20.5 C. Bradford: Sources of information on specific subject, Engineering **137**, 3550 (1934)
20.6 E. Garfield: Citation index for science, Science **122**, 108 (1955)
20.7 E. Garfield: Citation indexes in sociological and historical research, Am. Doc. **14**, 289–291 (1963)
20.8 J. Qiu, Y. Duan, J. Chen, E. Song, L. Ji: Review and prospect of the development of bibliometrics in China, Sci. Res. **21**(2), 143–148 (2003)
20.9 S. Luo: *Introduction to Bibliometrics* (Bibliography Literature Publishing House, Beijing 1986)
20.10 Y. Pan, Z. Ma: *Chinese S&T Journal Citation Reports (Core Edition) 2014* (Science and Technology Documentation, Beijing 2014)
20.11 J. Li, C. Cai: Analysis of the evaluation methods of various core periodicals, Chin. J. Sci. Tech. Period. **17**, 253–256 (2006)
20.12 Y. Pan, Z. Ma: *2003 Chinese S&T Journal Citation Reports (Core Edition)* (Scientific and Technical Documentation, Beijing 2003)
20.13 E. Garfield: Citation analysis as a tool in journal evaluation, Science **178**(4060), 471–479 (1972)
20.14 L. Jia: Determination of core periodicals and journal of chinese core periodicals, Inn. Mong. Univ. J. Philos. Soc. Sci. Ed. **6**, 119–124 (1997)
20.15 C. Li: *Comprehensive Evaluation of Academic Periodicals*, Dissertation (Central China Normal Univ., Changsha 2011)
20.16 W. Yue, S. Wilson: Measuring the citation impact of the journals in clinical neurology a structural equation modeling, Scientometrics **3**, 317–334 (2004)
20.17 X. Su: Construction of the evaluation system of academic periodicals of humanities and Social Sciences, Dongyue Luncong. **1**, 35–42 (2008)
20.18 J. Qiu: Study on indicator system and quantitative method of periodical evaluation, Mod. Libr. Inf. Technol. **7**, 23–26 (2004)
20.19 Y. Pan, Y. Zhang: Study on the indicator system of comprehensive evaluation of Chinese academic periodicals, Chin. J. Sci. Tech. Period. **11**, 217–219 (2000)
20.20 S. Luo: The origin and core elements of journal assessment, Libr. Inf. **S1**, 92 (1987)
20.21 L. Qin: VE selection method of core periodicals, Lib. Inf. Work **6**, 77 (1989)
20.22 Y. Wang: A new method for the determination of core periodicals – principal component analysis, J. China Soc. S&T Inf. **5**, 77–80 (1998)
20.23 S. Lin, R. Zeng: Emergence of journal evaluation and its development in China, Info. Res. **5**, 43–46 (2013)
20.24 D. Li: *Comparative Study of Chinese Periodical Evaluation System*, Disseration (Hebei Univ., Shijiazhuang 2013)
20.25 S. Zhuang: *Journal Overview of Chinese Core Periodicals* (Peking Univ. Press, Beijing 1992)
20.26 C. Wang: *Introduction to Bibliometrics* (Guangxi Normal Univ. Press, Nanning 1997)
20.27 H. Zhang: Analysis of citation times of Chinese management periodicals – A statistical analysis based on CSSCI 2004–2006, S&T Prog. Policy **26**, 21–22 (2006)
20.28 J. Qiu, C. Li: Analysis of academic influence of library and information science periodicals in China, Libraries **1**, 64–65 (2010)
20.29 X. Jin: Analysis of the influential factors of sci periodicals, Inf. Sci. **19**(10), 1055–1056 (2001)
20.30 X. Qin: *Study on Influential Factors of Academic Periodicals*, Disseration (Southwest Jiaotong Univ., Chengdu 2009)
20.31 M.E.D. Koenig: Bibliometric indicators versus expert opinion in assessing research performance, J. Am. Soc. Inf. Sci. **34**(2), 136–145 (1983)
20.32 Y. Wu: Never ignore the phenomenon of "Sleeping Beauty" in the literature, Science and Technology Times (Februray 29 2018)
20.33 W. Wang: Interdisciplinary evaluation: A reasonable path for the evaluation of comprehensive academic periodicals – A case study of the "Famous journal Project" of the Ministry of Education as a comprehensive journal, J. Nanjing Univ. Philos. Humanit. Soc. Sci. **48**(3), 140–146 (2011)
20.34 L. Li: Reflections on perfecting, the evaluation indicator system of Chinese STM periodicals, Acta Ed. **16**(4), 310–312 (2004)
20.35 Y. Huang, J. Wang, J. Kou: The relationship between social benefit and economic benefit of STM periodicals, Chin. J. Sci. Tech. Period. **10**(2), 105–106 (1999)
20.36 S. Ren, Z. Liu: Research on comprehensive evaluation method TS.74 of STM periodicals-principal component analysis, J. Chongqing Univ. Soc. Sci. Ed. **16**(1), 119–123 (2010)
20.37 M. Liu, Z. Ma, Y. Pan, L. Zhang, Z. Wang: Establishment of normalization model for quantitatively evaluating impact power of university's academic journals, Acta Ed. **16**(6), 405–406 (2004)
20.38 D. Xin, G. Han: Factor analysis application in the comprehensive evaluation of sci-tech journals, Math. Stat. Manag. **33**(1), 116–121 (2014)
20.39 H. Yiao: Comprehensive evaluation of periodicals based on rank and ratio method, Chin. J. Sci. Tech. Period. **17**(2), 213–215 (2006)
20.40 L. Yu, Q. Zhang, A. Liu: Multi-attribute evaluation in different disciplines-a case study of the Journal of Mathematics, Physics and Biology, Libr. Inf. Work **58**(20), 100–105 (2014)
20.41 Institute of Scientific and Technical Information of China: *Citation Report of China Science and Technology Journal (Core Edition) 2015* (Scientific and Technical Documentation, Beijing 2015)
20.42 J. Pang, Y. Zhang, Z. Ma: The research on the comprehensive system of chinese scientific & technical journal appraising factors, Chin. J. Sci. Tech. Period. **11**(3), 1–3 (2000)
20.43 C. Luo, J. Gao, M. Ma, S. Xiong: Analysis of the changes of compendex source periodicals of China mainland in 2013, Sci. Tech. Publ. **33**(2), 101–105 (2014)

20.44 The Science and Technology Daily: Reading "The International Influence promotion plan of STM periodicals" – An interview with Shen Eiming, a member of the Chinese Association of Science Associations, http://www.qstheory.cn/kj/kyzt/201212/t20121204_198242.htm (2012)
20.45 National Press and publication SAPPRFT: The 2014 National News Publishing industry basic situation, http://www.gapp.gov.cn/govpublic/80/965.shtml (2015)
20.46 X. Liu, Y. Yao: Statistics and analysis of STM periodicals in colleges and universities, Chin. J. Sci. Tech. Period. **26**(1), 86–90 (2015)
20.47 Y. Pan, Y. Zhang, Z. Ma: An analysis of the comprehensive academic index of Chinese English STM periodicals, Chin. J. Sci. Tech. Period. **14**(6), 614–617 (2003)
20.48 S. Ren, G. Zu: Analysis of the current situation of Chinese STM periodicals, Acta Ed. **16**(4), 303–305 (2004)
20.49 Z. Ma, Z. Yu, Y. Pan: Statistical analysis of the publishing situation of Chinese English STM periodicals, Chin. J. Sci. Tech. Period. **25**(10), 1278–1279 (2014)
20.50 The Office of the International Influence Promotion Program of China Science and Technology Periodicals: Circular on the announcement of the second phase of the China Science and Technology Journal's international Influence Promotion program, http://www.cast.org.cn/n17040442/n17041423/n17052289/17332641.html (2016)
20.51 Institute of Scientific and Technical Information of China: *2014 China Science and Technology Papers Statistics and Analysis (Annual Research Report) "Analysis of the Distribution of Science and Technology Theses in China"* (Scientific and Technical Documentation, Beijing 2016)
20.52 Institute of Scientific and Technical Information of China: *Citation Report of China Science and Technology Journal 2016 (Core Edition)* (Scientific and Technical Documentation, Beijing 2016)
20.53 H.J. Störig, S. Lu: *A Brief History of the World Language*, 2nd edn. (Shandong Pictorial Publishing House, Jinan 2009)
20.54 A. Swaan: *The Language of the World–The Global Language System* (Huacheng Publishing House, Guangzhou 2008)
20.55 S. Wang, H. Wang, R. Weldon: Bibliometric analysis of the English academic journals of China and their internationalization, Scientometrics **73**(3), 331–343 (2007)
20.56 Z. Ma, H. Tong, J. Yuan, C. Su: Quantitative analysis of the development and evolution of international papers on Chinese arts and humanities, J. Chin. Soc. Sci. Tech. Inf. **29**(5), 920–930 (2010)
20.57 W. Xu: *Research on Evaluation Indicator System of Chinese English STM Periodicals*, Disseration (Institute of Scientific and Technical Information of China, Beijing 2010)
20.58 Q. Chen: An analysis of the development of English STM periodicals in China, Chin. J. Sci. Tech. Period. **13**(5), 402–404 (2002)
20.59 S. Li: How to accelerate the internationalization of English STM periodicals, J. Hubei Univ. Econ. Humanit. Soc. Sci. Ed. **7**(5), 35–36 (2010)
20.60 Y. Chen: Discussion on the environment of Chinese English STM periodicals, Chin. J. Sci. Tech. Period. **15**(5), 593–595 (2004)
20.61 Y. Zhang, S. Wang, C. Lin: The development history and internationalization of English academic periodicals in China, Chin. J. Sci. Tech. Period. **14**, 754–757 (2003)
20.62 Z. Li, Y. Shi: Effective communication – An important measure of STM periodicals facing international – A case study of Chinese journal of materials science included in SCI, Chin. J. Sci. Tech. Period. **20**(5), 903–906 (2009)
20.63 B. Jin: On the development strategy of Chinese STM periodicals–on the internationalization of periodicals and the two development strategies for the world in periodicals, J. Shanxi Univ. Philos. Soc. Sci. Ed. **30**(3), 110–117 (2007)
20.64 W. Wu, X. Zhang, L. Zhou: The internationalization of Chinese STM periodicals, Chin. J. Sci. Tech. Period. **16**(4), 443–447 (2005)
20.65 J. Weng: The influence of Chinese STM periodicals in English, J. Jiangxi Libr. **36**(3), 67–68 (2006)
20.66 Z. Ma, Y. Pan: Peer review effect index of STM periodicals (EPR), J. Chin. Soc. Sci. Tech. Inf. **31**(3), 303–308 (2002)
20.67 J. Nortier: European language policy: Why is it so difficult?, http://www.slidefinder.net/e/european_language_policy_why_difficult/melbourneeuropeanlanguagepolicy/10323357 (2012)
20.68 Z. Ma, N. Wang, G. Zhou, Y. Pan: The analysis of mutual citation network on patterns of Chinese S&T core journal groups, Stud. Sci. Sci. **7**, 983–991 (2012)
20.69 Thomson Reuters: Journal Citation Reports, http://admin-apps.webofknowledge.com/JCR/JCR?RQ=HOME (2010)
20.70 Institute of Scientific and Technical Information of China: *Citation Report of China Science and Technology Journal 2011 (Core Edition)* (Scientific and Technical Documentation Press, Beijing 2011)
20.71 Y. Shi: Bibliometric analysis of Chinese medical journals of biomedical sciences, Chin. J. Sci. Tech. Period. **21**(6), 797–801 (2010)
20.72 Y. Guo, Z. Ma: Publishing state of SCI's Chinese academic journals cooperated with international publishers, Acta Ed. **23**, 555–557 (2011)
20.73 M. Zitt, E. Bassecoulard: Internationalization of scientific journals: A measurement based on publication and citation scope, Scientometrics **41**, 255–271 (1998)
20.74 Y. Pan, Z. Ma: *Chinese S&T Papers Statistics and Analysis 2010 Annual Research Report* (Scientific and Technical Documentation, Beijing 2012)
20.75 Z. Ma, Y. Pan: Effectiveness of peer review (EPR): A new indicator of academic journal, J. China Soc. Sci. Tech. Inf. **31**, 303–308 (2012)
20.76 Y. Pan, Z. Ma: *Chinese S&T Journals Citation Reports* (Scientific and Technical Documentation, Beijing 2011)

20.77 Z. Ma, N. Wang: The analysis of mutual citation network on patterns of Chinese S&T core journal groups, Stud. Sci. Sci. **30**, 983–991 (2012)
20.78 W. Xu: *Research on Evaluation Indicator System of Chinese English STM Periodicals*, Dissertation (Institute of Scientific and Technical Information of China, Beijing 2010)
20.79 Thomson Reuters: Journal Citation Reports, http://admin-apps.webofknowledge.com/JCR/JCR?RQ=HOME (2011)
20.80 Y. Yuan: *Argon atom model: A model structure for communication online*, Disseration (Wuhan Univ. 2006)
20.81 S. Ren, R. Ronad: International visibility of Chinese scientific journals, Scientometrics **53**, 389–405 (2002)
20.82 M. Zhang, X. Wu: The statement of Japan-made English journals, Chin. J. Sci. Tech. Period. **23**, 337–341 (2012)

Zheng Ma

Institute of Scientific and Technical Information of China (ISTIC)
Beijing, China
mazheng@istic.ac.cn

Zheng Ma is Professor at the Institute of Scientific and Technical Information of China. He is Deputy Director of the Center for Science and Technology Metrics. He graduated from Tsinghua University in 1999 with a major in environmental engineering. He is an executive member of the councils of the China Editology Society of Science Periodicals and the China Periodicals Association.

21. Bibliometric Studies on Gender Disparities in Science

Gali Halevi

Understanding gender related disparities in science is an essential step in tackling these issues. Through the years, bibliometric studies have designed several methodologies to analyze scholarly output and demonstrate that there are significant gaps between men and women in the scientific arena. However, gender identification in itself is an enormous challenge, since bibliographic data does not reveal it. These bibliometric studies not only focused on publication output and impact, but also on cross-referencing output, promotions and tenure data, and other related curriculum vitae (CV) information. This chapter discusses the challenges of tracking gender disparities in science through bibliometrics and reviews the various approaches taken by bibliometricians to identify gender and analyze the bibliographic data in order to point to gender disparities in science.

21.1 Background

According to the 2015 report of the *National Science Foundation* (NSF) on Women, Minorities, and Persons with Disabilities in Science and Engineering [21.1], there appears to be an inherit disproportion in the number of women holding academic positions as full professors. Although the number of female scientists has more than doubled since 1993, women currently occupy only about one-fourth of senior faculty positions. The main reason for this disparity seems to be the fact that there are older cohorts in the current science, engineering, and health workforces who are predominantly male.

In the past decade there have been global efforts to promote women's participation in science, technology, engineering, and mathematics (STEM). These efforts include educational programming from K-12 through higher education. Examples include government sponsored North American programs such as *Educate to Innovate* [21.2], the hEr VOLUTION, the Society for Canadian Women in Science and Technology and the

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_21

Canadian Coalition of Women in Engineering, Science, Trades, and Technology (CCWESTT). In Asia, there are several programs sponsored by not-for-profit organizations such as Destination Imagination in Singapore, the National Association of Women Entrepreneurs of Malaysia, as well as the Girls Code program in India, which encourages girls to participate in computer sciences and mathematics. The European Union formed *The European Centre for Women and Technology*, a partnership of more than 130 organizations that support women in technology from all over Europe. In Australia, the government has made a significant push to get more women involved in STEM education and careers and has pledged to invest $13 million over the next 5 years to make it happen. Finally, in developing countries there are a few organizations that are dedicated to training young women in STEM fields. These include the Organization for Women in Science in the Developing World, which creates research training and networking opportunities for female scientists, and the Elsevier Foundation, an organization that provides grants to support scholars in the early stages of their careers, to name a few.

Tracking the participatory levels of young women in STEM fields at the early years of their education is relatively well documented. For example, using high school registrations and completion of advanced placement (AP) examinations in chemistry, statistics, and calculus, policy makers in the United States were able to determine the participation percentages of young girls in STEM related fields [21.3]. In a similar manner, the NSF report on Women, Minorities, and Persons with Disabilities in Science and Engineering uses college degrees received to track academic degrees earned by women through the different STEM fields.

In Europe, the annual *She Figures* indicators report uses Eurostat (the Statistical Office of the European Union (EU)), which provides sex-disaggregated data on education, research and development, professional earnings and scientific employment, as well as primary data (broken down by gender) on senior academic staff, university leaders, funding applicants, and beneficiaries, as well as membership on boards of national research organizations ([21.4], Google search, no date).

United Nations Educational, Scientific, and Cultural Organization's (UNESCO's) report Girls and Women in Science, Technology, Engineering, and Mathematics in Asia ([21.5], Google search, no date), uses national examination results, as well as educational achievements for countries participating in the Program for International Student Assessment (PISA). The Trends in International Mathematics and Science Study (TIMSS) uses existing structures and results of assessments in mathematics and science at the national level to determine girls' participation levels in STEM academics.

These types of data, made available by national educational systems, enable researchers to gain insights into the state of female participation in different STEM fields from early school years through higher education. In addition, there are several studies that address not only the statistical levels of participation of women in STEM, but also attempt to explain the reasons for their underrepresentation. These psychological, educational, and sociological studies [21.6–9] have identified several influencing factors that contribute to the underrepresentation of women in STEM. Some of the factors identified are negative ability stereotypes and perceived bias, lack of role models, insufficient early experience, lack of peer support, and negative attitudes.

Although these studies might explain some of the barriers to female participation in STEM from early schooling through higher education by reporting on employment rates in research and academic institutions [21.1], they are not able to draw a holistic picture of female participation in science and are geographically and politically dependent.

In attempting to create a more comprehensive picture of the underrepresentation of women in STEM, bibliometric studies become an essential tool for tracking not only research participation, but also its consequential impact on scientific discovery. One of the manners in which participation and impact in scientific disciplines is measured is the number of publications published and the number of citations these receive. There are numerous statistically generated scores such as the h-index, field-weighted citation impact (FWCI), CiteScore, scholarly and social activity online, media mentions, and others, that attempt to measure the scientific impact of a scientist.

With these challenges in mind, the bibliometric community developed a variety of methodologies to be able to identify female participation and impact in the scientific arena. This chapter will review the methodologies used by bibliometricians to estimate female scientific participation and impact. This review is organized based on the purpose of the research in order to identify best practices for each. In this arena, both traditional metrics and altmetrics are used to measure the participatory levels and impact of women in science. Traditional metrics typically include the number of publications and the impact factor of the journals in which they are published, while the number of citations measures the impact of these articles. Altmetrics measures the social impact of the articles by examining engagement via social networks such as Twitter, LinkedIn, Facebook, and others. In addition, altmetrics measure the level to which these articles are used. This includes downloads, views, shares, and others. The use of traditional metrics and altmetrics combined allows researchers to measure both the number of articles contributed by women and their impact on the scientific endeavor.

In addition to these counts, bibliometric methods have been used to delve deeper into the participatory and impact levels of women in science. As will be shown in this chapter, several studies examined impact by analyzing the placement of the author in the authors list. First and last authors are sometimes an indication of the level of leadership in the production of the publication [21.10, 11]. In the area of citation counting, research examined not only the mere number of citations as a measure of impact but also the gender of the citing author/s. These studies focus on the level of peer recognition of female authors. The literature shows that there is a link between gender and citation rates wherein females are cited less than their male peers [21.12, 13].

21.2 Gender Determination

Bibliometric measures use publication information such as title, author name/s and affiliation/s to identify the contributors and assign the number of citations, mentions, and activities around the publication to them. When it comes to the ability to track and measure the relative contribution of female versus male authors and their respective impact, the main challenge is that gender is difficult to determine, since the only indicator for gender is the author's name. Although there are some names that can be determined as female or male, such as `Barbara` or `John`, there are numerous names that are gender neutral. An example is `Andrea`, which is a typically a female name in north America but a male name in Italy. In addition, there are gender-neutral names such as `Robin` or `Leslie`, which make the process of gender identification by name almost impossible just by reading them. There are names that can be difficult to identify as female or male because of regional traits. Such names include, for example, Asian, Middle Eastern, and African names, which are challenging to identify without deep research into each one. Finally, there is an additional complexity relating to issues of gender self-identification. Some authors might have been born as males or females and carry names that express their birth gender but self-identify differently. This is a rather recent social development, which thus far has not been investigated in gender-related studies.

Therefore, before discussing the bibliometric approaches to gender analysis in science, it is vital to review the main techniques used to determine the gender of the authors:

21.2.1 Manual Assignment

Gender verification is processed through the examination of the author's first name and, in cases where the name is difficult to verify, the author engages in a manual internet search for that person. Manual analysis is difficult to scale up or apply to large datasets. The largest dataset to be decoded for gender manually used 1 059 939 articles [21.14]. This study analyzed the contribution of female Russian scientists to the overall Russian body of scientific literature. Although the gender decoding was performed manually, the authors were able to determine gender in a straightforward way because Russian names are gender driven, making them easier to identify. Similarly, gender specific names assisted in the identification of female authors in Poland [21.15]. Needless to say, such an approach would be difficult to apply to author names that do not adhere to traditional gender driven rules.

21.2.2 Institutional Rosters and National Databases

Studies focusing on institutional or national levels assume that larger amounts of articles produced by female authors can serve as an indication of their productivity and participatory levels. These studies decode gender by using existing databases that include gender information [21.16–19]. National databases can include data of federal or state agencies, which contain educational statistics and include the gender of the scientists. Examples are The Spanish Council for Scientific Research, The American Doctoral Dissertations Database, The Indian Directory of CSIR Scientists, and so forth. In these studies, the data collection includes the retrieval of articles assigned to an institution or a nation and cross-checking the names with national databases that hold the gender of the authors.

21.2.3 Questionnaires

A few studies included questionnaires in order to verify authors' gender and seniority, in addition to using institutional records. This was done mostly when the study spanned several universities across various locations or when a specific institution was examined [21.11–13, 20–22].

Table 21.1 Gender identifying software tools and databases

Name of software or database	Website
Gender Checker	http://www.genderchecker.com/search.aspx
Genderize.io	https://genderize.io/
Gender Guesser	http://www.genderguesser.com/
NamePedia	http://www.namepedia.org/
Name database	https://github.com/organisciak/names
Name database	http://www.ssa.gov/oact/babynames/
Chinese Name Gender Guesser	
Social Security US	https://www.ssa.gov/oact/babynames/
Gender API	https://www.gender-api.com/
First Name Sex	http://www.firstnamesex.com/

21.2.4 Software Tools

Finally, some studies on participation and productivity measures used specifically developed software and web tools that assist in identifying the gender of the author [21.23, 24]. These software tools rely on name data collected from sources such as name registries or national databases such as social security indexes to create matching algorithms that can identify whether a name is female or male (Table 21.1).

The accuracy of these approaches varies. Manual identification of gender by author names reports between 10 and 15% exclusion of articles due to the inability to identify the gender of the authors. Studies using software tools compliment the initial identification of the author names by manually searching for author names that the software was not able to identify. Studies using university rosters and national databases report the most accurate results with relatively large data samples [21.25–32]. However, it should be noted that no one methodology is superior. When the purpose of the study is to examine female participation in specific areas of science, the publication-level approach can serve as a good indicator. Choosing high impact journals or conferences and extracting their publications will ensure that the study focuses on a well-defined discipline or sub-discipline. However, in studies that span nations and institutions around the world, it is more efficient to obtain data from local rosters or national databases in a comprehensive manner. Studies in this arena could also benefit from using existing software tools for initial name screening and gender identification and compliment data gathering with manual searches on the internet. Although not very sophisticated, there are several name registries that are available freely on the web and provide gender identification of names in different regions of the world.

21.3 Definitions

The studies examining gender disparities in science while using bibliometric methodologies can be grouped into four main areas:

1. Productivity
2. Performance
3. Impact and visibility
4. Academic standing.

In the following sections of this chapter each of these areas of study will be discussed while covering the three major elements of the research investigation:

1. Unit of investigation
2. Data collection and dataset used
3. Research methodology.

21.3.1 Areas of Investigation

In order to better understand both the areas of investigation and the elements within the bibliometric approach there is a need to define them so that they can be further applied for future studies.

Productivity

In gender studies, the number of publications one produces mostly defines productivity. These include journal articles, conference papers, books, and book chapters. These publications are peer reviewed and indexed in either a controlled database such as the Web of Science or Scopus or in a non-controlled database such as Google Scholar. The main premise is that the more publications one can attribute to an author, the more productive the author is. Therefore, productivity studies rely on publication count per author.

Performance

The manner by which bibliometrics defines performance is mostly through applying analytical calculations to the number of publications and the number of citations they receive. The assumption is that the more citations an author's publications receive, the higher his/her performance. The impact factor (IF) of the journal in which they are published plays a major role. The higher the IF score of the journal in which the author publishes, the higher his/her performance is considered to be. Measures such as h-index scores, which use the number of citations and number of publications, are used in this context. Bibliometrics also developed several citations-driven analytics to publication data that aim to capture performance within the context of each scientific field and drive a more accurate capture of an author's performance. The field weighted citation impact (FWCI) measure, for example, aims to measure the impact of publications and citations within each field and is used to capture performance.

Impact

Impact is defined as both the number of times a publication is cited and the overall engagement an author's publications receive. Engagement is usually measured by altmetrics, which count the number of downloads, views, shares, and mentions of publications in academic and social networks. The underlying reason for such measurement is the assumption that publications have a wider range of impact, which goes beyond citations. Many publications might be read, shared, and discussed but not necessarily officially cited. Such engagement is considered impact and is attributed to the author/s whether or not their work is cited.

Career Track—Tenure and Promotions

In this arena, research is focused on examining the progression and development of authors' academic careers. Using bibliometric data and deep analytics of the authors' CVs for academic career progress, these studies aim to establish whether there is a link between productivity and impact and academic retention, promotion, and recruitment.

21.4 Research Approach

Units of investigation: Gender studies are seen to take two main approaches: (1) publication level or (2) department, institution, or country level. These require different data collection approaches and cross-analytics.

21.4.1 Publication Level Analysis

The journal/article level approach includes studies that collect data for a specific journal or proceedings over a span of time and analyze the number of female and male authors in each article or conference paper [21.25, 26, 28, 29, 31–33]. Interestingly, the vast majority of these studies are in the area of medicine, covering clinical practices such as surgery, emergency medicine, psychiatry, dermatology, and others [21.10, 23, 26, 31, 34–43]. There are very few publication-level studies that look at the output rates of females versus males that tackle specific areas of science [21.11, 23, 32]. The data collection described in these studies is straightforward. The procedures employed in these studies usually include the selection of 3–4 high-impact journals based on impact factor scores or the analysis of a specific journal over a few years.

21.4.2 Institutional and/or National Level

In most cases, these studies are used to determine the productivity and participation level of females in science overall rather than specific disciplines. Large areas such as STEM are usually addressed on a national or institutional level while using large amounts of publications verified for gender by national databases [21.16, 44, 45]. Institutional level studies look at the productivity and participation rates of female scientists across scientific areas in a specific academic institution [21.17, 46]. These studies utilize the institutional roster as a way to identify the gender of the authors. In many cases, studies are conducted on current faculty members and compare publications rates over time based on the number of authored papers and an author's seniority.

21.5 Data Collection and Datasets Used

Number of publications: The data is collected by retrieving the publications of each author, institution, department, or state, depending on the purpose of the research. In institution, department, or state level studies, the investigators retrieve the publications from known databases such as the Web of Science, Scopus, or Microsoft Academic and use the affiliation name and address to identify them. In studies that examine a known group of scientists, investigators use the authors' names to retrieve their publications. It should be noted here that each database could potentially retrieve different numbers of publications per author due to varying coverage policies. Unlike Google Scholar, controlled databases such as the Web of Science and Scopus have a selective approach to indexing coverage. Therefore, in many cases, the complete corpus of one's publications will not be retrieved. The data source in this case is crucial. Depending on the discipline, one should consider which database to use [21.47].

21.5.1 Number of Citations

Citations are counted per publication retrieved. Citation counts also depend on the database used. In some studies, citations are also calculated and normalized for a discipline, institution, or country [21.15, 17, 20, 48, 49]. The main purpose of calculating normalized citations is to achieve the outmost accuracy when comparing different disciplines, institutions, departments, etc. Normalization accounts for the size of the institution or department and the overall citations rates of a scientific field. In this manner, larger institutions or scientific fields are not compared to smaller ones without accounting for these differences. Some fields cite in higher rated publications than others, in the same way that some institutions have more faculty members than others [21.50–52]. When comparing citation rates, these should be taken into consideration as well.

21.5.2 Authors

Gender studies are focused on authors. The datasets selected cover different groups of authors. These datasets vary from one study to another and include the use of questionnaires distributed within an institution or department. The use of national databases include:

1. Spain: The Council for Scientific Research
2. Spain: CSIC database [21.29]
3. USA: American Doctoral Dissertations [21.53]
4. India: Bibliography of Doctoral Dissertations
5. India: Directory of CSIR scientists [21.19]
6. France: Centre National de la Recherche Scientifique (CNRS) [21.54]
7. Italy: Italian Observatory of Public Research [21.17]
8. USA: The US Social Security Database [21.55, 56]
9. USA: US AMA Fellowship and Residency Interactive Database [21.37, 57]
10. USA: US National Center for Education Statistics (NCES) [21.18]; use of institutional databases (rosters) [21.58]
11. South Africa: ten universities [21.59].

21.5.3 Publication Datasets

Publications are selected based mostly on their impact factor and within their disciplines in order to measure female participation in these fields [21.10, 23, 24, 31, 34, 35, 37, 40, 43, 60–62]. Interestingly, many of these studies focus on different medical fields while using a collection of leading journals to retrieve publications by female authors. There are some, although not many, that focus on other fields of science such as neuroscience, astronomy, library and information science, international relations, materials science, and computer science [21.21, 25, 26, 28, 29, 32, 63–65].

21.6 Methodology

21.6.1 Counting

This method uses simple counts to compare the number of female authors to the number of male authors [21.10, 14, 29, 31, 33–35, 37, 39, 41, 43, 62, 65]. This method is frequently used when the study focuses on a selection of journals or conference proceedings and compares the number of female authors to male authors over time. Many of these studies focus on high impact journals or conference proceedings in different disciplines.

21.6.2 Normalizing

This method is frequently used in the area of performance comparisons between female and male authors across disciplines. Disciplinary examination of performance requires careful consideration of the publication rates of different disciplines, as well as their citations and even grant amounts. Therefore, these studies must normalize the results to account for differences in the number of publications, citations, and even grants [21.16, 17, 30, 48, 66–68].

21.6.3 Matching and Clustering

This method uses matching and clustering techniques to compare not only gender but also chronological age, academic age, experience, and other qualifying variables [21.68, 69]. Studies using this technique tend to look at a specific population in an institution or a geographic location and conduct analysis of sub-groups based on the above characteristics. Since there are a number of qualifiers that are considered per gender group, these studies tend to be small but accurate as far as the analysis is concerned. Many of the studies using these methods focus on analyzing output rates between smaller sub-groups of the population studied [21.70–74]. A good example of using a variety of matching techniques is a study researching Danish health sciences graduates [21.75]. In this study, the authors studied 541 students enrolled at the Institute of Clinical Research at the University of Southern Denmark. The relatively small population studied allowed the authors to match the population studied based on their discipline and sub-discipline of study, their age, education, and time lapse from the enrolment date. This multimatching approach ensures that there is as little bias as possible in the analysis. However, as with other studies in this area, an approach like this requires that a relatively small population be studied and a reliable source of data, in this case, the university records.

21.6.4 Qualitative Measures via Surveys and Interviews

Qualitative methods are used especially in studies that look at career development differences between females and males. In order to track career paths and development, these studies use surveys and interviews to gain insights into the differences between males and females [21.69].

21.7 Productivity

Studies in the area of productivity mostly compare female and male scientific output in a variety of disciplines and regions of the world while sketching overall trends in female participation in science compared to that of males. These studies define *productivity* as the total number of publications produced by a scientist. This simple yet effective approach to such comparisons aims to demonstrate whether or not there is a gap between the numbers of publications produced by females as compared to those produced by males [21.26, 29, 37, 38, 42, 45, 55, 62, 76].

The majority of the studies examining productivity in the form of the number of article contributions by female scientists use the numbers as an indication of participation in a certain field. The main assumption is that an increase in female authors is an indication of an overall increase of female participation in the field. This could explain why many of these studies are in the area of medicine, a field that is still considered male dominated.

Despite the fact that this could be perceived as a simplistic approach to measure productivity, it is an effective way to examine male/female participation in specific areas of science. The main reason for choosing this approach is that it enables a straightforward selection of topics, body of literature, and time frames for analysis. For example, a study on the productivity rates of males and females in general medicine will require a selection of journals in the discipline, retrieval of publications in specific time frames, and an analysis of the number of publications produced by females as compared to those produced by males. In the same manner, a study can also examine whether females occupy the positions of first authors as opposed to second or third authors. By selecting a well-defined collection of journals, researchers can also examine the participation of females in science over time. This is especially true for well-established journals that have been publishing for decades, such as the *New England Journal of Medicine* or the *Journal of the American Medical Association*.

Interestingly, these type of studies are seen to be very popular in various clinical research areas [21.10, 23, 31, 32, 35, 37, 39, 41–43, 60, 61]. By selecting a few high impact journals and comparing female authorship to male authorship, researchers have been able to demonstrate an overall steady increase of female-authored papers in clinical research fields such as plastic surgery, pediatrics, ophthalmology, oncology, radiology, and others. In addition to clinical sciences, productivity in terms of number of publications has also been studied in the areas of library and information science (LIS), material sciences, computer science, nanoscience, and neuroscience [21.25, 26, 28, 29, 32, 33, 49, 65]. One of the few social sciences-related studies focused on the differences in productivity between female and male researchers in the areas of sociology and linguistics [21.67].

Productivity and output rates have also been used to measure the overall participation of females in science on a national level. In order to analyze the productivity rates of females and males on a national level, studies have been designed using two main approaches: (1) aggregating the literature to the institutional and national levels and (2) utilizing national databases and institutional rosters to identify female authors across departments and institutions and aggregating them to a national level [21.14, 15, 17, 18, 45, 46, 48, 56, 77–79].

Aggregation of the literature to a national level is often conducted when there is an interest in examining female participation in a specific area of science in a specific country. In these cases, the studies select a defined collection of publications in the discipline and limit it by year ranges. By using affiliations' names and locations, researchers retrieve the publications that list the country. A dataset containing publications, which lists the country, is then limited by year range and analyzed for female/male names. The results are usually able to show not only female–male participation rates in a scientific area, but also the quality of the journals in which they publish. In addition, these types of studies can usually serve as indicators of a country's level of publishing in high-quality publications.

There are several challenges with using this approach. First, the body of literature must be carefully selected. In order to measure national levels of female/-male authorships, one must include international and national publications in order to create a well-balanced dataset. This is especially true in disciplines such as arts and humanities and social sciences, which normally have areas of research that are unique to a country. These could include linguistics, social work, law, and others. Second, when retrieving publications based on names of affiliations, the dataset will not account for an author's publications published under an affiliation prior to the one they currently hold. For example, if a study is looking to examine the rates of Spanish female/male publication rates in pharmacology between 2000 and 2016, a list of publications that cover pharmacology will be defined. Once the list is defined, all publications that list Spain in the affiliation will be retrieved. Once the dataset is retrieved, the author names will be analyzed in order to find the number of females versus males. However, the names of the researchers in the dataset will probably not capture the entirety of the female or male scientist's body of work, as some could have moved outside of Spain in these years. This could be problematic when aggregating to a national level. Therefore, studies compare the results with current rosters or databases available via the institution or federal agencies or validate the author's affiliation through a CV.

Another approach is using national or institutional databases that contain names of scientists to identify their publications. In most cases, the names will be searched for in combination with affiliation name and countries in order to avoid including erroneous records. This approach works most effectively when the dataset is focused on specific areas of science. By using this approach, researchers are able to identify female–male participation ranges in specific areas of research. The challenge in this approach is the number of publications analyzed. Conducting large-scale studies over time and across disciplines requires high computational capabilities, which are not readily available at many research institutions.

Overall, studies found that despite of the steady increase of female-authored publications, there is still a gender gap when it comes to the number of publications produced by females and males in most areas of science and regardless of geography. In an attempt to explain this phenomenon, *Cole* and *Zuckerman* [21.53] listed four explanations of what they termed *the productivity puzzle*. These include:

1. Scientific ability
2. Self-selection
3. Social selection
4. Accumulated disadvantage.

Despite no previous findings pointing to the fact that gender is a factor of scientific ability, *Cole* and *Zuckerman* claim that various biological and psychological factors influence scientific output and put females at an inherent disadvantage. In addition, life events, which *Cole* and *Zuckerman* define as *self-selection* such as child bearing and family obligations, are seen to interrupt women's scientific careers and, therefore, their overall output rates. This was also found in other studies showing that the women begin their careers later than men due to life events and thus produce less publications in their early career [21.69, 80]. Late career mainly means that Ph.D. and Post doc training periods usually coincide with women's marriage and child bearing periods. This, in turn, creates a reality in which women are seen to produce less publications than men due to pregnancies and maternity leave, which put them at a disadvantage when returning to the workplace [21.81]. Trying to catch up with their male counterparts after periods of low productivity is very difficult.

The notion of *social-selection*, as explained by *Cole* and *Zuckerman*, pertains to how scientific productivity of female researchers is sometimes affected by social factors that could be discriminatory. Overall in society, men still dominate positions of power. This is also true in the scientific arena. Therefore, there are preexisting unfavorable conditions to the entry of females into research organizations, which drives lower productivity and participation in science [21.82–84].

There are several reasons identified in the literature that attempt to explain the gap in productivity between females and males, as show below.

21.7.1 Underrepresentation

Females are still underrepresented in many areas of science. Despite the growing numbers of females in academia, many scientific fields are still male dominated [21.39, 56, 85, 86]. There are certain disciplines that are considered more *female orientated*, which include nursing; midwifery; speech, language, and hearing; education; social work; and librarianship. Areas such as military sciences, engineering, robotics, aeronautics, physics, computer science, life sciences, and medicine are still dominated by men. Therefore, it is understandable that the number of publications attributed to males in these disciplines will be greater than those attributed to females.

21.7.2 Career Development

Females begin their careers later than males do. The career timeline of men and women is different. During their Ph.D. and post-doctoral years, men and women are seen to embark on different career paths [21.87, 88]. While men are seen to produce publications at these times, women are diverted by marriage and starting a family. These life events are seen to affect women's productivity more than men's. Time taken for maternity leave and child care creates a gap in productivity that can be difficult to bridge. Once women return to full academic and research activity they are at a disadvantage and need to make up for the time dedicated to family, for instance. This sociological issue could potentially explain the productivity gap seen in early career stages. However, it does not provide a full explanation for the publication gap in later years [21.69, 89–93].

21.7.3 Specialization Versus Diversification

The publication gap seems to persist later in women's careers due to several academic and scientific tendencies that have been identified in the literature. Studies have found that specialization promotes productivity. This mainly applies to increasing proficiency, reputation, and expertise building. Women seem to specialize less, which affects their productivity. There is an interesting difference between men's and women's perception of specialization; while men are of the overall opinion that specialization promotes excellence, women think that diversification in research activities will broaden their professional networks and the breadth of their scholarly activity [21.67, 94, 95].

21.7.4 Collaboration and Professional Networks

Men and women differ in their approaches to professional collaboration and networking. Women tend to achieve better performance in domestic and/or small collaborative groups. Men tend to create wider collaborative networks that overlap with their own, while women focus on smaller and homogenous networks [21.66, 69, 96, 97]. Yet despite of their overall research diversification tendencies, women do not follow this pattern when it comes to research collaboration, which is seen to be smaller and more focused.

21.7.5 Research Versus Teaching

Women are seen to focus on teaching and service rather than research [21.18, 98, 99]. There is a still a debate as to whether the focus on research is a choice or a result of bias. However, this phenomenon is prevalent across institutions and disciplines, which again can explain the gap in productivity rates between men and women.

21.8 Research Performance

Studies in this area use bibliometric measures such as citation rates, h-index scores [21.100], and research collaboration network analysis [21.30, 58, 66, 72] to determine the level of performance of scientists and compare males to females.

Within the area of performance studies, researchers have looked at gender differences in citation rates. Analyzing the number of citations that female-authored publications receive and comparing those to male-authored publications is a predominant method used in these studies [21.12, 13, 21, 22, 64, 101]. There are no definitive conclusions when it comes to citations rates. Some studies find that there are no major differences between males and females with regards to citation rates [21.46, 53, 86, 102, 103], while others find that female-authored publications receive more citations than male-authored ones [21.81, 104], and still others find that female-authored publications receive fewer citations than male-authored ones [21.56, 105].

The inconclusiveness of the results is an indication of the complexity of the subject. On the one hand, the ability to track citations to female-authored papers is straightforward, yet the explanation as to why is more complex. In disciplines where there are fewer female-authored publications, the low number of citations of their work is easily explained. However, in disciplines where the female–male ratio is more balanced, an explanation is difficult. Therefore, these results must be put in the context of their disciplines and even within the specialities they examine [21.96, 106, 107]. There is some evidence that female authors tend to be more selective in their choice of publications, while males are seen to produce as many publications as possible, most of which do not receive any citations. Females, on the other hand, might publish less but their articles receive higher rates of citation [21.108].

Adding to the level of complexity are studies that examine the male to female citation rates. These studies reveal the phenomenon that males do not cite women at the same rate that they cite their male counterparts. The reasons behind this are not clear. Some explanations include, again, disciplinary differences, networking within the research community, visibility, and even the type of publication as factors influencing the male–female citations rates. In areas of research that are inherently male dominated, higher male–male citations are to be expected because the number of male authors is larger. In addition, research has shown that there is an unconscious bias in citation patterns, as well as closely tight networks that generate a high degree of male-to-male citations [21.109].

Studies have also examined whether there is an inherent bias in the peer-review process that could influence female citation performance [21.110] and found no evidence of this. This is also true with regards to funding. Research has shown that the success rate of women and men are virtually equivalent when it comes to National Institutions Of Health (NIH) grant funding (for example, [21.36]). The main reason for the performance gap is the overall lower percentages of women's participation in science overall. This study found that large numbers of women prefer to pursue more independent career paths that are not strongly tied to NIH funding and make this decision earlier in their careers when they are seen to focus on family and children. Otherwise, men and women are more likely to be equally funded throughout their careers.

21.9 Impact and Visibility

Unlike performance measures that can be tracked by productivity and citations, impact and visibility are becoming broader in scope. With the advent of technology, there are several new metrics that aim at capturing user interaction with scientific literature. A *user*, in this broader context, applies not only to the professional reader, but also to the general public. The assumption in measuring these interactions is that a scientific publication can have an impact beyond citations, especially when relevant to the general public. Measures, such as the number of downloads, clicks, reads, mentions in social media channels, journalistic articles, and others can be counted towards impact. Citations, on the other hand, are selective. While writing an article one might read more articles than one cites. These read articles should be counted towards an author's overall impact. The same applies to views of articles, tables and graphs, mentions and discussions on social media channels, and so forth. These metrics generally are referred to as *altmetrics*.

There is growing evidence that scholars are adopting social media as an instrument to communicate, network, and promote their research [21.111–116]. Participation of academics in social networking ranges from using LinkedIn, ResearchGate, and other professional networking channels, as well as Twitter, Facebook, YouTube, and others. Thus far, there is very little research to be found in the area of gender and altmetrics.

One study that measured altmetrics as a function of impact in gender studies uses http://altmetrics.com as a platform to track impact [21.117]. The findings reveal that social media is more gender-balanced than citations, explaining that 'the scientific community which constitute the citing audience is more male-dominated than the social media environment' [21.117, p. 42]. A study focusing on the number of papers and readership rates of women and men on Mendeley [21.118] also shows that readership is more gender balanced.

Although we cannot conclude that impact is more gender balanced when measured through altmetrics, there is initial evidence that supports this. Using social media channels to promote one's work and achievements could be a path for women to become more visible.

An interesting study looking at the impact of web presence on productivity and citations [21.65] used a combined methodological approach to analyze male and female scientists in nanoscience and nanotechnology. The initial objectives of the study were to examine whether or not having a web presence correlates with a higher or lower number of publications and citations. The manner in which the authors approached such an analysis was to select first authors in defined journals in nanoscience and nanotechnology within specific year ranges. Once the list of authors was extracted, the researchers manually searched for each of the authors using Google in order to determine their gender as well as their web presence outlets. *Web presence* was defined as any online channel featuring that author. This could have included official university profiles, blogs, personal web pages, and others. The authors searched for over 1000 names in Google in order to identify gender and web presence.

Interestingly, this study found that the number of publications does not have a direct correlation with web presence. This means that there is no significant difference between the number of publications produced by web-present scholars and web-absent ones. However, with regards to citations, web-present scholars are shown to receive more citations than web-absent ones. Males and females are equally represented online in these areas of research, and females are equally as recognized (e. g., cited) as their male counterparts. Although this study is focused on two disciplines, its methodology can be easily applied to others. Overall, it seems that women can use social networks and social media to their advantage. Whereas previous research showed that women tend to have rather small and closed networks, having a strong web presence on professional social networking platforms could serve women in expanding their networks and reaching a wider audience, thus driving attention to their work and achievements.

Networking has a direct impact on visibility as well as on impact [21.119]. Being visible and active in the research community drives impact as people read and are aware of one's research. This seems to hold true regardless of gender. Studies in the area of collaborative patterns are not restricted to bibliometric investigations and include business, economics, and social explorations of the networking patterns of men and women in different work environments [21.72, 115, 120–122]. Bibliometric studies in this area, however, focus on the examination of whether men and women differ in their collaboration and networking practices, which, in turn, can influence visibility, productivity, and impact [21.66, 69, 123, 124]. The overall conclusions of these studies indicate that men and women have different networking behaviors and motivations [21.120, 121, 125]. Women, for example, join networks in order to develop their skills and seek mentors, but mostly for reasons beyond career motivation, including support and informal interactions. Men, on the other hand, are motivated more by career development and enhancement. This is not surprising considering the fact that there are still networks that are not accessible to women, including senior level management groups that operate on the premise of the *old boys' network* of men who know each other and share similar hobbies or recreational preferences.

A study by *Lee* and *Bozeman* [21.124] used a survey of approximately 400 scientists affiliated with research centers in the USA in order to discover whether age, gender, rank, or marital status influence productivity and collaboration. The survey found that people mostly collaborate with others within their local or immediate settings and environments. This was found to be true regardless of gender or other social factors. Collaboration was also found to have a positive impact on the number of publications produced. The 2005 initial dataset in *Bozeman* and *Gaughan* [21.123] was based on a national database. This set contained the names of the individual researchers working in multidisciplinary work groups or research areas, especially in centers funded by the National Science Foundation (NSF) and by the Department of Energy (DoE). Once the gender of the individuals had been established, the study used a questionnaire that included questions about research collaboration, grants and contracts, job selection, work environment, and demographic information. Although the study initially hypothesized that females would have less collaborators due to social dynamics, the results showed that, in fact, the number of collaborators is more a function of tenure and grant size. Therefore, if the principal investigator (PI) is a female awarded a large grant, her collaborative network will be as cosmopolitan and as large as that of a male under the same circumstances. That said, the study did find that female scientists have a higher percentage of female collaborators compared to males, especially when it comes to non-tenured females. However, tenured female faculty members collaborate with females at the same rate as males. Therefore, one can conclude that non-tenured females will seek to collaborate more with females rather than males. This was somewhat explained by the mentorship role that females tend to take.

The methodologies used in these studies rely heavily on combined data, which include individual names, gender, and CVs. Therefore, they are limited in scope and tend to target specific groups. Despite of the accuracy of the results, and their ability to portray a given situation, they are very difficult if not impossible to scale, since they rely on surveys and interviews.

21.10 Careers: Recruitment and Promotions

One of the most essential areas in the study of the gender gap in academia is the topic of recruitment and promotions, as they pertain to women's careers. There are several studies that attempt to demonstrate female rank and career development by looking at different academic institutions and disciplines. Tracking an academic career is a challenging task, since it requires not only identifying the gender of the authors but also detailed time-dependent information on their professional ranks and career progression. Most journals do not include an author's academic rank, which makes it almost impossible to track career development via bibliographic information. This is the main reason why this topic has been examined in bibliometrics by harnessing samples of publications combined with either interviews, CVs, or data from national databases depicting recruitment and tenure procedures of state universities [21.58, 71, 74, 76, 126].

The most common methodology used in order to track women's academic careers uses a collection of journals in a defined discipline and year ranges. Once the articles from these journals have been identified and downloaded, the researchers identify the gender of the authors either manually or by combining name check software tools. The next step for tracking the development of careers, tenures, and promotions is to send a survey and conduct interviews with as many authors as possible. These surveys and interviews provide insight into issues of time lapses between degrees obtained, promotions, tenures, and academic roles.

Another methodology seen in this arena is the use of national and/or professional databases to identify female academics. For example, *Hancock* et al. [21.71] utilized data from membership in the International Studies Association (ISA), and *Abramo* et al. [21.126] used publicly available data on the 2008 Italian Competition for Associate Professors, which includes data on the winners as well as their gender. Once the population had been identified, a bibliographic search for publications and citations was conducted. Both these approaches resulted in datasets that are a combination of qualitative and quantitative measures.

When tracking careers, the most common areas researchers look into are comparisons of the careers of women and men, as they pertain to three major areas:

1. Probability of gaining promotions and/or tenure
2. Probability of holding major leadership roles within academic institutions
3. Salary equality wherein women's salaries are compared to those of their male counterparts in the same positions. This data is mostly collected via interviews and surveys.

In order to discover biases, most of the studies utilize bibliometric data, such as number of publications and citations. The main premise is that if a female author has similar amount of publications and citations to a male in the same discipline, their rank, salary, and roles should be equal.

Research results show that the geographic location and type of academic institution play a role in achieving equality. In Italy, for example, *Abramo* et al. [21.126] found that in certain disciplines where there is a high number of males compared to females, males are most likely to win an academic position even if they are less deserving of the position than a female counterpart. In addition, there seems to also be favoritism, as a male committee president is more likely to award an academic position to a male candidate whom he has known for several years, again, regardless of merit. In The Netherlands, *van den Besselaar* and *Sandström* [21.74] found that productivity differences between men and women play a role in career development. In early career years, men and women show similar levels of performance. This changes in later years, when women diverge into teaching positions, while men choose research-oriented positions, which in turn affects their respective ranks and roles. The authors hypothesize that gender bias could play a role in this self-selection process, as women are seen to choose less competitive positions due to a lack of support compared to their male counterparts. In Sweden, on the other hand, *Mählck* [21.58] found that females do not perceive gender to be a factor in their academic career development despite the differences in rank between them and their male counterparts. The authors in this case assume that the overall culture of equality in the country is a factor influencing this perception, even if gaps exist.

The gender gap in career development is, therefore, found to be dependent on geographic location, culture, and self-selection, as well as, in some cases, favoritism and discrimination. In some cultures, females are discouraged from pursuing scientific careers and are either self-diverted or encouraged to pursue teaching rather than research paths. Methodologically, data collection for studies on this topic is complex and involves manual procedures and qualitative applications of surveys, for example.

21.11 Summary

Gender studies are seen to focus on topics including productivity, impact, visibility, performance, and career development. Methodologies in these areas include, in most cases, more than one procedure, because of the complex data collection. Table 21.2 summarizes the major approaches to the different studies and is organized on the basis of the four main comparative objectives of gender studies:

1. Productivity
2. Performance
3. Impact and visibility
4. Career development.

Each topic, except career tracking, presents some similarities in data collection approaches that usually begin by identifying high impact factor journals in a discipline. In most cases, the article collection is limited by year range. This approach is one of the most popular ones seen in the literature; it allows for a close study of a specific discipline by targeting journals that are considered core journals. The proportions of female and male authors in these journal publications provide an indication of impact measures as well as of productivity over time. Publishing overall is an indication of productivity, while the IF of the journals indicates impact. Comparing datasets of articles from different years allows researchers to notice which gender group publishes more and in what types of journals, as well as to track female participation levels over time.

Once the dataset of articles has been downloaded, the gender of the authors needs to be identified. This portion of the data collection presents a challenge. Unlike other bibliographic metadata, gender is not provided in the articles. Therefore, gender identification relies on the first name of the author. As was seen above, there are a few techniques to do so. The use of national name databases is helpful. These databases are usually made available by governments and list the most common names and their gender. They are, of course, not exhaustive, but they do provide some basic information about the gender of names in a specific country. The second technique is to use automated software

Table 21.2 Summary of common data collection approaches

Comparative measure	Most common data collection approaches
General female/male comparisons using bibliometric analysis	Journal selection: High impact factor journals in specific discipline/s are selected. Year range selection: Dependent on the purpose of study. If comparing output overtime articles are retrieved in various years or year ranges. Article retrieval: Articles are retrieved using an indexing database. Bibliographic data must include first names of author/s. Identifying gender: Female/male gender identification is done either manually, using a software tool, utilizing national or institutional databases, and a combination of the above. When gender cannot be identified, some studies utilize questionnaires or direct communications with the authors
Productivity: Most commonly measures the number of publications output	The number of articles attributed to female authors versus male authors is compared
Performance: Mostly measures citations received and grants received.	For article citations: Data is collected as above. Citations per article are collected using a citation tracking database such as Scopus or the Web of Science. For grants: If the study focuses on a specific institution and/or country, grant information is collected either through an open state or federal database or through the institution's own records. In cases where the study examines a specific discipline in various geographical locations, data is collected through article retrieval, identification of leading authors, followed by grant information retrieval.
Impact and visibility: Mostly measures the overall awareness of publications and authors.	Web presence: Tracked by searching for authors online, mostly manually. Web presence is attributed when an author has a website, blog, or any professional network activity. Altmetrics: Article level metrics such as usage, social media mentions, media, and others are collected via different platforms such as http://altmetric.com, Mendeley, and others
Career: Tenure and promotion	Institutional or national databases containing promotions and/or tenured appointments are used to compare female and male career progression. Surveys: Individual surveys to designated populations are used to collect information about promotions and tenure. CV analysis: This is used in many cases when an institution or department is studied. There are cases where CVs are available via state or federal institutions, which allows for a larger population to be studied

that allows for name searches and identifies their gender. These software tools are helpful but do not always properly identify gender. A third technique is manual checking of each name using the internet and finding information such as pictures or websites to identify the gender of authors. Finally, departmental or institutional human resources records are also used. This approach is, of course, the most accurate but limits the study and does not allow conclusions to be made beyond the institution. In general, any of the above techniques will always present a challenge if there is an attempt to scale it up. Since there is still no fully automated way to identify the gender of the authors, manual work will always be required.

In order to measure performance, researchers use the number of publications and the number of citations per article. This is the most common indicator of high performance. While these measures are quite straightforward, impact and visibility are more complex measures to analyze. In this arena, researchers use altmetrics indicators such as usage, media mentions, and others to measure the overall impact of articles published by women versus those published by men. Visibility was seen to be measured by web presence such as participation in online professional networks and having a blog or other online website.

Career development is more complex to track. In order to be able to measure promotions, tenure, and career progression of men and women, most studies use CV analysis combined with national or institutional records that provide gender information as well as career progression over time.

References

21.1 National Science Foundation: Women minorities and persons with disabilities in science and engineering, https://www.nsf.gov/statistics/women (2017)

21.2 The White House President Barack Obama: Education for K-12 students, https://obamawhitehouse.archives.gov/issues/education/k-12/educate-innovate

21.3 College Board: Education Pays – Trends in Higher Education (2013)

21.4 European Commission: She Figures, https://ec.europa.eu/research/swafs/pdf/pub_gender_equality/she_figures_2015-final.pdf (2015)

21.5 United Nation Educational Scientific and Cultural Organisation: Girls and women in science, technology, engineering and mathematics in Asia, http://unesdoc.unesco.org/images/0023/002315/231519e.pdf (2015)

21.6 J. Saucerman, K. Vasquez: Psychological barriers to STEM participation for women over the course of development, Adultspan J. **13**(1), 46–64 (2014)

21.7 C.A. Shapiro, L.J. Sax: Major selection and persistence for women in STEM, New Dir. Inst. Res. **2011**(152), 5–18 (2011)

21.8 M.-T. Wang, J. Degol: Motivational pathways to STEM career choices: Using expectancy-value perspective to understand individual and gender differences in STEM fields, Dev. Rev. **33**(4), 304–340 (2013)

21.9 M. Adya, K.M. Kaiser: Factors influencing girls' choice of information technology careers. In: *Encyclopedia of Gender and Information Technology*, ed. by E.M. Trauth (IGI Global, London 2006)

21.10 G. Filardo, B. da Graca, D.M. Sass, B.D. Pollock, E.B. Smith, M.A.-M. Martinez: Trends and comparison of female first authorship in high impact medical journals: Observational study (1994–2014), BMJ **352**, i847 (2016)

21.11 R. Sidhu, P. Rajashekhar, V.L. Lavin, J. Parry, J. Attwood, A. Holdcroft, D.S. Sanders: The gender imbalance in academic medicine: A study of female authorship in the United Kingdom, J. R. Soc. Med. **102**(8), 337–342 (2009)

21.12 D. Maliniak, R.M. Powers, B.F. Walter: The gender citation gap. In: *APSA 2013 Annu. Meet. Paper* (2013), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2303311

21.13 L. Turner, J. Mairesse: *Individual Productivity Differences in Public Research: How Important are Non-Individual Determinants? An Econometric Study of French Physicists' Publications and Citations (1986–1997)* (Centre National de La Recherche Scientifique, Paris 2005), http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.535.9659&rep=rep1&type=pdf

21.14 A. Paul-Hus, R.L. Bouvier, C. Ni, C.R. Sugimoto, V. Pislyakov, V. Larivière: Forty years of gender disparities in Russian science: A historical bibliometric analysis, Scientometrics **102**(2), 1541–1553 (2014)

21.15 B.M. Webster: Polish women in science: A bibliometric analysis of polish science and its publications, 1980–1999, Res. Eval. **10**(3), 185–194 (2001)

21.16 G. Abramo, C.A. D'Angelo: Accounting for gender research performance differences in ranking universities, Curr. Sci. **109**(10), 1783–1789 (2015)

21.17 G. Abramo, C.A. D'Angelo, A. Caprasecca: Gender differences in research productivity: A bibliometric analysis of the Italian academic system, Scientometrics **79**(3), 517–539 (2009)

21.18 J.P. Gander: Faculty gender effects on academic research and teaching, Res. High. Educ. **40**(2), 171–184 (1999)

21.19 B.M. Gupta, S. Kumar, B.S. Aggarwal: A comparision of productivity of male and female scientists of CSIR, Scientometrics **45**(2), 269–289 (1999)

21.20 M.A. Ferber: Citations and networking, Gend. Soc. **2**(1), 82–89 (1988)

21.21 B. McElhinny, M. Hols, J. Holtzkener, S. Unger, C. Hicks: Gender, publication and citation in sociolinguistics and linguistic anthropology: The construction of a scholarly canon, Lang. Soc. **32**(3), 299–328 (2003)

21.22 S.M. Mitchell, S. Lange, H. Brus: Gendered citation patterns in international relations journals, Int. Stud. Perspect. **14**(4), 485–492 (2013)

21.23 M. Mimouni, O. Segal, S. Zayit-Soudry, A.Y. Nemet, Y. Barak, N. Geffen: Female authorship in major academic ophthalmology journals: Trends in recent years, Investig. Ophthalmol. Vis. Sci. **57**(12), 5558–5558 (2016)

21.24 C.L. Piper, J.R. Scheel, C.I. Lee, H.P. Forman: Gender trends in radiology authorship: A 35-year analysis, Am. J. Roentgenol. **206**(1), 3–7 (2016)

21.25 M.I. Atinmo, S.W. Jimba: Gender and authorship patterns in an African librarianship journal, 1991–1997, Libr. Rev. **51**(9), 458–463 (2002)

21.26 D. Dubey, A. Sawhney, A. Atluru, A. Amritphale, A. Dubey, J. Trivedi: Trends in authorship based on gender and nationality in published neuroscience literature, Neurol. India **64**(1), 97 (2016)

21.27 M.E. Falagas, E.I. Pitsouni, G.A. Malietzis, G. Pappas: Comparison of PubMed, Scopus, Web of Science, and Google scholar: Strengths and weaknesses, FASEB Journal **22**(2), 338–342 (2008)

21.28 S. Gul, T.A. Shah, S.N. Hamade, R. Mushtaq, I. Koul: Effects of gender in library and information science research: A case study of The Electronic Library, Electron. Libr. **34**(3), 488–503 (2016)

21.29 E. Mauleón, M. Bordons: Productivity, impact and publication habits by gender in the area of materials science, Scientometrics **66**(1), 199–218 (2006)

21.30 M.W. Nielsen: Gender consequences of a national performance-based funding model: New pieces in an old puzzle, Stud. High. Educ. **42**(6), 1–23 (2015)

21.31 E. Nkenke, R. Seemann, E. Vairaktaris, H.-G. Schaller, M. Rohde, F. Stelzle, C. Knipfer: Gender trends in authorship in oral and maxillofacial surgery literature: A 30-year analysis, J. Cranio-Maxillofac. Surg. **43**(6), 913–917 (2015)

21.32 G.C. Vilanilam, H.V. Easwer, S. Vimala, A. Radhakrishnan, B.I. Devi, S.N. Nair: Women and neuroscience publishing: Is the gender gap closing in?, Neurol. India **64**(3), 583 (2016)

21.33 J.M. Cohoon, S. Nigai, J.J. Kaye: Gender and computing conference papers, Communications ACM **54**(8), 72–80 (2011)

21.34 A.A. Ahmed, B. Egleston, E. Holliday, G. Eastwick, C. Takita, R. Jagsi: Gender trends in radiation oncology in the United States: A 30-year analysis, Int. J. Radiat. Oncol. Biol. Phys. **88**(1), 33–38 (2014)

21.35 V. Franco-Cardenas, J. Rosenberg, A. Ramirez, J. Lin, I. Tsui: Decadelong profile of women in ophthalmic publications, JAMA Ophthalmology **133**(3), 255–259 (2015)

21.36 T.J. Ley, B.H. Hamilton: The gender gap in NIH grant applications, Science **322**(5907), 1472–1474 (2008)

21.37 A.M. Paik, L.J. Mady, N.L. Villanueva, E. Goljo, P.F. Svider, F. Ciminello, J.A. Eloy: Research productivity and gender disparities: A look at academic plastic surgery, J. Surg. Educ. **71**(4), 593–600 (2014)

21.38 A. Raj, P.L. Carr, S.E. Kaplan, N. Terrin, J.L. Breeze, K.M. Freund: Longitudinal analysis of gender differences in academic productivity among medical faculty across 24 medical schools in the United States, Acad. Med.: J. Assoc. Am. Med. Coll. **91**(8), 1074–1079 (2016)

21.39 M.S. Sharkey, R.S. Feinn, V.V. Tate, C.W. Carter, T.T. Cassese: Disproportionate participation of males and females in academic pediatric orthopaedics: An analysis of abstract authorship at POSNA 2009–2013, J. Pediatr. Orthop. **36**(4), 433–436 (2016)

21.40 D. Sing, D. Ouyang, S.S. Hu: Gender trends in authorship of spine-related academic literature: A 38-year perspective, Spine J. **16**(10), S279–S280 (2016)

21.41 J.F. Waljee: Discussion: Gender authorship trends of plastic surgery research in the United States, Plastic Reconstruct. Surg. **138**(1), 143e–144e (2016)

21.42 E.J. Yun, D.Y. Yoon, B. Kim, J.Y. Moon, S.J. Yoon, S.J. Hong, S. Baek: Closing the gender gap: Increased female authorship in AJR and radiology, Am. J. Roentgenol. **205**(2), 237–241 (2015)

21.43 S. Zhang, H.-Y. Kim, R.E. Hill, E. Veledar, S.C. Chen: A 10-year comparison of women authorship in US dermatology literature, 1999 vs. 2009, Int. J. Women's Dermatol. **2**(1), 1–4 (2016)

21.44 D.W. Aksnes, K. Rorstad, F. Piro, G. Sivertsen: Are female researchers less cited? A large-scale study of Norwegian scientists, J. Am. Soc. Inform. Sci. Technol. **62**(4), 628–636 (2011)

21.45 M.R. Davarpanah, H.M. Moghadam: The contribution of women in Iranian scholarly publication, Libr. Rev. **61**(4), 261–271 (2012)

21.46 E. Mauleón, C. Daraio, M. Bordons: Exploring gender differences in patenting in Spain, Res. Eva. **23**(1), 62–78 (2014)

21.47 G. Halevi, H. Moed, J. Bar-Ilan: Suitability of Google scholar as a source of scientific information and as a source of data for scientific evaluation—Review of the literature, J. Informetr. **11**(3), 823–834 (2017)

21.48 M. Bordons, F. Morillo, M.T. Fernández, I. Gómez: One step further in the production of bibliometric indicators at the micro level: Differences by gender and professional category of scientists, Scientometrics **57**(2), 159–173 (2003)

21.49 K. Goel: Gender differences in publication productivity in psychology in India, Scientometrics **55**(2), 243–258 (2002)

21.50 L. Bornmann, R. Haunschild: Relative citation ratio (RCR): An empirical attempt to study a new field-normalized bibliometric indicator, J. Assoc. Inform. Sci. Technol. **68**(4), 1064–1067 (2017)

21.51 L. Bornmann, A. Thor, W. Marx, H. Schier: The application of bibliometrics to research evaluation in the humanities and social sciences: An exploratory study using normalized Google scholar data for the publications of a research institute, J. Assoc. Inform. Sci. Technol. **67**(11), 2778–2789 (2016)

21.52 L. Leydesdorff, T. Opthof: Normalization at the field level: Fractional counting of citations, http://arxiv.org/abs/1006.2896 (2010)

21.53 J.R. Cole, H. Zuckerman: The productivity puzzle. In: *Advances in Motivation and Achievement. Women in Science* (JAI, Greenwich 1984)

21.54 J. Mairesse, L. Turner: A look at individual differences in scientific research productivity: An econometric analysis of the publications of the French CNRS physicists in condensed matter (1980–1997). In: *Conf. Rethinking Sci. Policy: Anal. Frameworks Evidence-Based Policy, SPRU, Brighton* (2002) pp. 21–23

21.55 J.D. West, J. Jacquet, M.M. King, S.J. Correll, C.T. Bergstrom: The role of gender in scholarly authorship, PloS One **8**(7), e66212 (2013)

21.56 V. Larivière, C. Ni, Y. Gingras, B. Cronin, C.R. Sugimoto: Global gender disparities in science, Nature **504**(7479), 211–213 (2013)

21.57 J.A. Eloy, P.F. Svider, D.V. Cherla, L. Diaz, O. Kovalerchik, K.M. Mauro, S.S. Chandrasekhar: Gender disparities in research productivity among 9952 academic physicians, Laryngoscope **123**(8), 1865–1875 (2013)

21.58 P. Mählck: Mapping gender differences in scientific careers in social and bibliometric space, Sci. Technol. Hum. Values **26**(2), 167–190 (2001)

21.59 D. Jacobs: A bibliometric study of the publication patterns of scientists in South Africa 1992–96, http://www.informationr.net/ir/6-3/paper104.html (2016)

21.60 M. Amering, B. Schrank, I. Sibitz: The gender gap in high-impact psychiatry journals, Acad. Med. **86**(8), 946–952 (2011)

21.61 J.D. Feramisco, J.J. Leitenberger, S.I. Redfern, A. Bian, X.-J. Xie, J.S. Resneck: A gender gap in the dermatology literature? Cross-sectional analysis of manuscript authorship trends in dermatology journals during 3 decades, J. Am. Acad. Dermatol. **60**(1), 63–69 (2009)

21.62 R. Jagsi, E.A. Guancial, C.C. Worobey, L.E. Henault, Y. Chang, R. Starr, E.M. Hylek: The "gender gap" in authorship of academic medical literature—A 35-year perspective, N. Engl. J. Med. **355**(3), 281–287 (2006)

21.63 J.M. Cavero, B. Vela, P. Cáceres, C. Cuesta, A. Sierra-Alonso: The evolution of female authorship in computing research, Scientometrics **103**(1), 85–100 (2015)

21.64 M. Håkanson: The impact of gender on citations: An analysis of college and research libraries, Journal of Academic Librarianship and Library Quarterly, Coll. Res. Libr. **66**(4), 312–323 (2005)

21.65 H. Sotudeh, N. Khoshian: Gender differences in science: The case of scientific productivity in nano science and technology during 2005–2007, Scientometrics **98**(1), 457–472 (2014)

21.66 G. Abramo, C.A. D'Angelo, G. Murgia: Gender differences in research collaboration, J. Informetr. **7**(4), 811–822 (2013)

21.67 E. Leahey: Gender differences in productivity research specialization as a missing link, Gend. Soc. **20**(6), 754–780 (2006)

21.68 H. Prozesky, N. Boshoff: Bibliometrics as a tool for measuring gender-specific research performance: An example from South African invasion ecology, Scientometrics **90**(2), 383–406 (2012)

21.69 S. Kyvik, M. Teigen: Child care, research collaboration, and gender differences in scientific productivity, Sci. Technol. Hum. Values **21**(1), 54–71 (1996)

21.70 G. Abramo, C.A. D'Angelo, F. Rosati: Selection committees for academic recruitment: Does gender matter?, Res. Eval. **24**(4), 392–404 (2015)

21.71 K.J. Hancock, M.A. Baum, M. Breuning: Women and pre-tenure scholarly productivity in international studies: An investigation into the leaky career pipeline, Int. Stud. Perspect. **14**(4), 507–527 (2013)

21.72 H. Kretschmer, I.F. Aguillo: New indicators for gender studies in web networks, Inf. Process. Manag. **41**(6), 1481–1494 (2005)

21.73 K. Rørstad, D.W. Aksnes: Publication rate expressed by age, gender and academic position—A large-scale analysis of Norwegian academic staff, J. Informetr. **9**(2), 317–333 (2015)

21.74 P. van den Besselaar, U. Sandström: Gender differences in research performance and its impact on careers: A longitudinal case study, Scientometrics **106**(1), 143–162 (2016)

21.75 T.F. Frandsen, R.H. Jacobsen, J.A. Wallin, K. Brixen, J. Ousager: Gender differences in scientific performance: A bibliometric matching analysis of Danish health sciences graduates, J. Informetr. **9**(4), 1007–1017 (2015)

21.76 H. Kretschmer, T. Kretschmer: Gender bias and explanation models for the phenomenon of women's discriminations in research careers, Scientometrics **97**(1), 25–36 (2013)

21.77 J. Requena, D. Vargas, C. Caputo: Gender in science in Venezuela: Fading of the gap, Interciencia **41**(3), 162–170 (2016)

21.78 T. Söderlund, G. Madison: Characteristics of gender studies publications: A bibliometric analysis based on a Swedish population database, Scientometrics **105**(3), 1347–1387 (2015)

21.79 B. Malacuso, V. Larivière, T. Sugimoto, C.R. Sugimoto: Is science built on the shoulders of women? A study of gender differences in contributorship, Acad. Med. **91**(8), 1136–1142 (2016)

21.80 S. Stack: Gender, children and research productivity, Res. High. Educ. **45**(8), 891–920 (2004)

21.81 M.R. Symonds, N.J. Gemmell, T.L. Braisher, K.L. Gorringe, M.A. Elgar: Gender differences in publication output: Towards an unbiased metric of research performance, PLoS One **1**(1), e127 (2006)

21.82 K. Prpić: Gender and productivity differentials in science, Scientometrics **55**(1), 27–58 (2002)

21.83 T.M. Timmers, T.M. Willemsen, K.G. Tijdens: Gender diversity policies in universities: A multi-perspective framework of policy measures, High. Educ. **59**(6), 719–735 (2010)
21.84 Y. Xie, K.A. Shauman: Sex differences in research productivity: New evidence about an old puzzle, Am. Sociol. Rev. **63**(6), 847–870 (1998)
21.85 G. Lewison, V. Markusova: Female researchers in Russia: Have they become more visible?, Scientometrics **89**(1), 139–152 (2011)
21.86 G. Lewison: The quantity and quality of female researchers: A bibliometric study of Iceland, Scientometrics **52**(1), 29–43 (2001)
21.87 L.J. Sax, L.S. Hagedorn, M. Arredondo, F.A. DiCrisi: Faculty research productivity: Exploring the role of gender and family-related factors, Res. High. Educ. **43**(4), 423–446 (2002)
21.88 K. Ward, L. Wolf-Wendel: Academic motherhood: Managing complex roles in research universities, Rev. High. Educ. **27**(2), 233–257 (2004)
21.89 M.F. Fox: Women, science, and academia graduate education and careers, Gend. Soc. **15**(5), 654–666 (2001)
21.90 M.F. Fox: Gender, family characteristics, and publication productivity among scientists, Soc. Stud. Sci. **35**(1), 131–150 (2005)
21.91 L. Grant, I. Kennelly, K.B. Ward: Revisiting the gender, marriage, and parenthood puzzle in scientific careers, Women's Stud. Q. **28**(1/2), 62–85 (2000)
21.92 S. Kyvik: Motherhood and scientific productivity, Soc. Stud. Sci. **20**(1), 149–160 (1990)
21.93 A.Y. Lewin, L. Duchan: Women in academia, Science **173**(4000), 892–895 (1971)
21.94 M.F. Fox, P.E. Stephan: Careers of young scientists: Preferences, prospects and realities by gender and field, Soc. Stud. Sci. **31**(1), 109–122 (2001)
21.95 B. Keith, J.S. Layne, N. Babchuk, K. Johnson: The context of scientific achievement: Sex status, organizational environments, and the timing of publication on scholarship outcomes, Soc. Forces **80**(4), 1253–1281 (2002)
21.96 N. Carayol, M. Matt: Individual and collective determinants of academic scientists' productivity, Inf. Econ. Policy **18**(1), 55–72 (2006)
21.97 L. Grant, K.B. Ward: Gender and publishing in sociology, Gend. Soc. **5**(2), 207–223 (1991)
21.98 J. Miller, M. Chamberlin: Women are teachers, men are professors: A study of student perceptions, Teach. Sociol. **28**(4), 283–298 (2000)
21.99 R.K. Toutkoushian, M.L. Bellas: Faculty time allocations and research productivity: Gender, race and family effects, Rev. High. Educ. **22**(4), 367–390 (1999)
21.100 M. Martinez, S. Lopez, K. Beebe: Gender comparison of scholarly production in the musculoskeletal tumor society using the Hirsch index, J. Surg. Educ. **72**(6), 1172–1178 (2015)
21.101 J.B. Slyder, B.R. Stein, B.S. Sams, D.M. Walker, B.J. Beale, J.J. Feldhaus, C.A. Copenheaver: Citation pattern and lifespan: A comparison of discipline, institution, and individual, Scientometrics **89**(3), 955–966 (2011)
21.102 C.A. Copenheaver, K. Goldbeck, P. Cherubini: Lack of gender bias in citation rates of publications by dendrochronologists: What is unique about this discipline?, Tree-Ring Res. **66**(2), 127–133 (2010)
21.103 A. Ledin, L. Bornmann, F. Gannon, G. Wallon: A persistent problem, EMBO Reports **8**(11), 982–987 (2007)
21.104 Á. Borrego, M. Barrios, A. Villarroya, C. Ollé: Scientific output and impact of postdoctoral scientists: A gender perspective, Scientometrics **83**(1), 93–101 (2010)
21.105 L. Hunter, E. Leahey: Parenting and research productivity: New evidence and methods, Soc. Stud. Sci. **40**(3), 433–451 (2010)
21.106 M. Leahy, J. Doughney: Women, work and preference formation: A critique of Catherine Hakim's preference theory, J. Bus. Syst. Gov. Ethics **1**(1), 37–48 (2006)
21.107 H.-M. Puuska: Effects of scholar's gender and professional position on publishing productivity in different publication types. Analysis of a Finnish university, Scientometrics **82**(2), 419–437 (2010)
21.108 U. Sandström: Combining curriculum vitae and bibliometric analysis: Mobility, gender and research performance, Res. Eval. **18**(2), 135–142 (2009)
21.109 I. Tahamtan, A.S. Afshar, K. Ahamdzadeh: Factors affecting number of citations: A comprehensive review of the literature, Scientometrics **107**(3), 1195–1225 (2016)
21.110 C.W. Fox, C.S. Burns, A.D. Muncy, J.A. Meyer: Gender differences in patterns of authorship do not affect peer review outcomes at an ecology journal, Funct. Ecol. **30**(1), 126–139 (2016)
21.111 A.M. Djuricich: Social media, evidence-based tweeting, and JCEHP, J. Contin. Educ. Health Prof. **34**(4), 202–204 (2014)
21.112 C. Guerin, S. Carter, C. Aitchison: Blogging as community of practice: Lessons for academic development?, Int. J. Acad. Dev. **20**(3), 212–223 (2015)
21.113 D. Lupton: "Feeling better connected": Academics' use of social media [Text], http://apo.org.au/node/53908 (2014)
21.114 M. McPherson, K. Budge, N. Lemon: New practices in doing academic development: Twitter as an informal learning space, Int. J. Acad. Dev. **20**(2), 126–136 (2015)
21.115 B. Stewart: Open to influence: What counts as academic influence in scholarly networked Twitter participation, Learn. Media Technol. **40**(3), 287–309 (2015)
21.116 M. Thelwall, K. Kousha: ResearchGate articles: Age, discipline, audience size, and impact, J. Assoc. Inf. Sci. Technol. (2016), https://doi.org/10.1002/asi.23675
21.117 A. Paul-Hus, C.R. Sugimoto, S. Haustein, V. Larivière: Is there a gender gap in social media metrics? In: *Proc. Int. Soc. Scientometr. Inform. Conf., Istanbul, Turkey* (2015)

21.118 J. Bar-Ilan, I. van der Weijden: *Altmetric Gender Bias?–Preliminary Results* (STI, Leiden 2014) p. 26

21.119 G. Abramo, C.A. D'Angelo, F. Di Costa: Research collaboration and productivity: Is there correlation?, High. Educ. **57**(2), 155–171 (2009)

21.120 Y. Benschop: The micro-politics of gendering in networking, Gend. Work Org. **16**(2), 217–237 (2009)

21.121 C. Cross, C. Armstrong: Understanding the role of networks in collective learning processes: The experiences of women, Adv. Dev. Hum. Resour. **10**(4), 600–613 (2008)

21.122 S. Durbin: Creating knowledge through networks: A gender perspective, Gend. Work Org. **18**(1), 90–112 (2011)

21.123 B. Bozeman, M. Gaughan: How do men and women differ in research collaborations? An analysis of the collaborative motives and strategies of academic researchers, Res. Policy **40**(10), 1393–1402 (2011)

21.124 S. Lee, B. Bozeman: The impact of research collaboration on scientific productivity, Soc. Stud. Sci. **35**(5), 673–702 (2005)

21.125 I.H. van Emmerik: Gender differences in the creation of different types of social capital: A multilevel study, Soc. Netw. **28**(1), 24–37 (2006)

21.126 G. Abramo, C.A. D'Angelo, F. Rosati: Gender bias in academic recruitment, Scientometrics **106**(1), 119–141 (2016)

Gali Halevi

Icahn School of Medicine at Mount Sinai
New York, NY, USA
gali.halevi@mssm.edu

Gali Halevi is the Associate Dean of Libraries and Information Sciences at the Icahn School of Medicine at Mount Sinai. Gali oversees several physical libraries and library services to all Mount Sinai affiliations. Prior to this she was part of a team at Elsevier developing research and evaluation metrics tools. Areas of interest include research evaluation and output, global scientific trends, and international research collaborations.

22. How Biomedical Research Can Inform Both Clinicians and the General Public

Elena Pallari, Grant Lewison

This study involved the collection of clinical practice guidelines (CPGs) on five noncommunicable disease (NCD) areas from 21 European countries, and extraction of their evidence base in the form of papers in journals processed on the Web of Science (WoS). We analyzed these cited papers to see how their geographical provenance compared with European research in the respective subjects and found that European research (and that from the USA, Australia, and New Zealand) was over-cited compared with that from East Asia. In cancer, surgery and radiotherapy research made important contributions to the CPGs.

We also collected medical research stories from 30 newspapers from 22 European countries and the WoS papers that they cited. There was a heavy emphasis on cancer, particularly breast cancer, and its epidemiology, genetics, and prognosis, but new treatment methods were seldomly reported, particularly surgery and radiotherapy. Some of the stories quoted commentators, with those from the two UK newspapers often mentioning medical research charities, which thereby gained much free publicity.

Both sets of cited research papers showed a marked tendency to be over-cited by documents from their countrymen; the ratio was higher the smaller the country's contribution to research in the subject area.

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_22

22.1 Study Objectives

The main purpose of biomedical research is to improve healthcare [22.1], both by the better treatment of patients and by the prevention of illness [22.2]. The second of these goals is often given lower priority by national healthcare systems, because the immediate need to treat patients claims more attention [22.2]. This is, of course, a common problem in policy-making, summed up neatly in the phrase, *we were fighting off the alligators, but the real need was to drain the swamp*. In effect, longer-term problems that could provide great benefit at a modest cost are neglected in favor of short-term problems that are crying out for a solution [22.3]. The question of what treatment to provide should be answered with reference to the best available science [22.4–6], but instead it is often based on the personal experience of clinicians [22.7, 8], and the lobbying of special interests [22.9], particularly in healthcare by pharmaceutical companies [22.10, 11].

22.1.1 Importance of Study

We considered that it was important to examine two ways in which biomedical research could influence these two goals. The first is to examine the evidence underlying clinical practice guidelines (CPGs), which are increasingly being used to determine patient treatment [22.12]. The second is to look at the stories in the mass media, which are the main means whereby research is brought to the attention of the public [22.13]. The *public* includes a wide range of people, from politicians who decide healthcare policy, their expert advisers, clinicians and other healthcare personnel, other researchers, and of course the general public. Nowadays, we are being encouraged to take a more active role in the protection of our own health, assisted by public health legislation [22.14, 15]. The latter depends to a large extent on public consensus, and good timing, so that it will readily be put into practice [22.16]. Examples of evidence-based policy [22.17] are the mandatory use of car seat belts [22.18], and the prohibitions on smoking in enclosed spaces such as offices and restaurants [22.19]. Others currently being considered, but having difficulty making headway against determined industrial lobbying, are minimum prices for alcoholic beverages and restrictions on sugary soft drinks [22.20].

The work to be described formed part of a major European Union (EU)-funded project on the mapping of European research on five noncommunicable diseases over 12 years, 2002–2013. *Europe* or EUR31 was defined as the 28 member states of the EU, plus Iceland, Norway, and Switzerland. The five NCDs were cardiovascular disease including stroke (cerebrovascular disease), designated as CARDI; diabetes or DIABE; mental disorders or MENTH; cancer or ONCOL; and respiratory diseases or RESPI. This was undertaken in 2014–2015 by King's College London (KCL) in association with six partners: the London School of Economics (LSE, the coordinator); in Estonia, the Estonian Research Council, Tartu; in France, Université Paris Est Créteil, Paris; in Germany, Technische Universität Berlin; in Italy, Università Commerciale Luigi Bocconi, Milan; and in Spain, Escuela Andaluza de Salud Pública, Granada. These partners assisted us with the updating and calibration of the five NCD filters that were used to identify research papers in the Web of Science (WoS, © Clarivate Analytics, formerly part of Thomson Reuters) and with the work described in this chapter. We also recruited KCL graduate students from most of the other EU Member States with the necessary research and language skills to read European CPGs and newspaper stories citing research.

22.1.2 The Development of Clinical Practice Guidelines in Europe

Clinical practice guidelines started to be described in the literature in 1971 [22.21], and the first papers were all from the USA. There was rather little notice taken of them in the 1970s and 1980s, but interest really began in the 1990s. After something of a lull in the 2000s, interest picked up in the 2010s (Fig. 22.1). There was also a shift in the countries that were involved. In the 1970s it was only the USA, but the UK started publishing in the 1980s, and was gradually overtaken by the other EUR31 countries, and the rest of the world (Fig. 22.2).

Figure 22.2 makes clear that the European Union (plus the two European Economic Area (EEA) states and Switzerland) has been increasingly dominant. In parallel with the growing numbers of research papers, there has been a similar growth in the numbers of CPGs published in European countries, see Fig. 22.3 which shows the numbers that concern diabetes.

Although our search for these CPGs was not exhaustive, we did identify ones from 21 countries out of the 31. In the UK, there are two organizations that publish CPGs: the Scottish Intercollegiate Guidelines Network (SIGN), which began operations in 1993 [22.22], six years before the one for England and Wales, the National Institute for Clinical and Care Excellence (NICE) [22.23]. In France, Guides Parcours de Soins are published by the Haute Autorité de Santé [22.24].

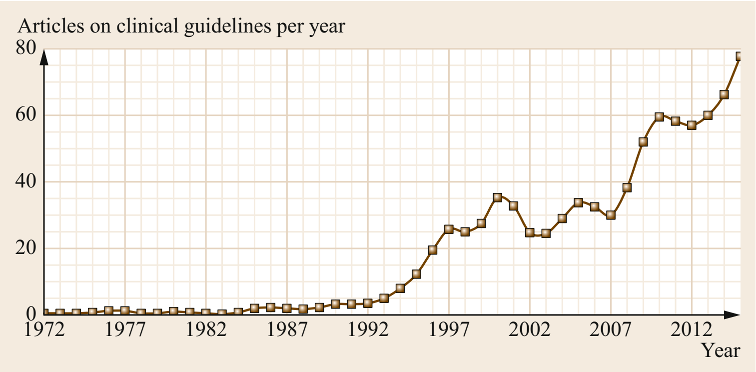


Fig. 22.1 Increase of numbers of articles in the Web of Science (WoS) with clinical guidelines in their title from 1971 to the present, 3-year running means

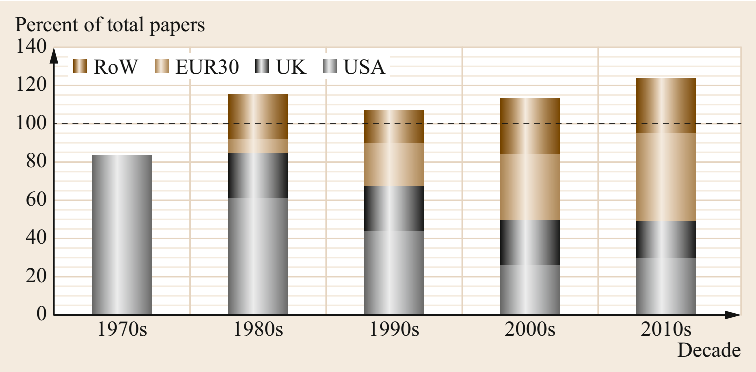


Fig. 22.2 Percentages of total set of clinical guidelines articles in the WoS in each of five decades from the USA, the UK, the rest of Europe (EUR30), and the rest of the world (RoW), integer counts. Note: percentages sum to more than 100% because of international collaboration. In the 1970s, some papers had no recorded addresses

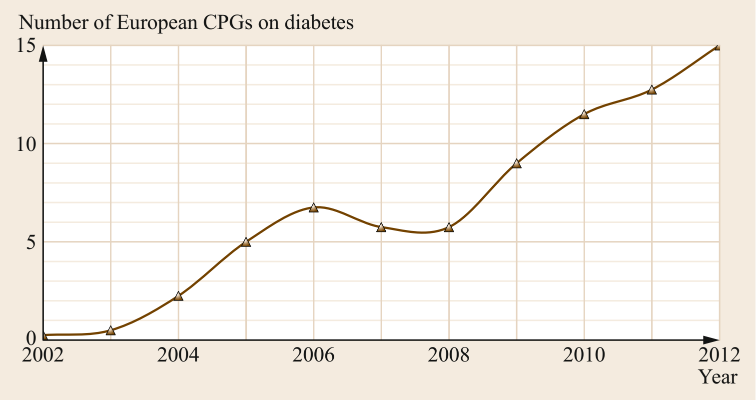


Fig. 22.3 Numbers of clinical practice guidelines in diabetes from EUR31 countries, 2002–2013, 3-year running means

In Germany, a large number of organizations, some federal and some private nonprofit, come together to prepare and publish CPGs, *Nationale VersorgungsLeitlinen* [22.25] as the collection of logos on a diabetes guideline shows (Fig. 22.4). In Italy, some CPGs are developed by the Istituto Superiore di Sanità in partnership with its parent department, the Ministero della Salute [22.26]; others are produced by nonprofit societies (Figs. 22.5 and 22.6). This is also common in Spain, where the societies form a group and the CPGs are published in an academic journal, but other guidelines are sponsored by national and regional ministries [22.27] (Figs. 22.7 and 22.8). So there is a wide variety of publishers of European CPGs, and both governments and nonprofit organizations are involved. A more comprehensive list of European CPG providers is provided in the Appendix, Table 22.17.

22.1.3 Previous Work on the References on Clinical Practice Guidelines

Although clinical practice guidelines started to be described in the literature in the 1970s, it was almost two decades before their impact on medical practice was

Fig. 22.4 The German organizations that were associated with a diabetes guideline

Fig. 22.5 The Italian organizations that were associated with a diabetes guideline

evaluated [22.28] or their evidence, in the form of the cited references, began to be considered as a means to evaluate biomedical research [22.29, 30]. Grant's conclusions are still valid, namely that the papers cited on CPGs published by the NICE are clinical rather than basic; that they are fairly recent; and that they tend to over-cite research by own-country authors. This last conclusion strictly only applied to British guidelines as there does not appear to be any comparable study based on those of other countries. Subsequently, the scope of these studies was extended to a wider selection of CPGs [22.31–33], including those from the SIGN and the British Medical Association's handbook, *Clinical Evidence*. Subsequently, *Kryl* et al. [22.34] showed that the references on two CPGs from the NICE, on dementia and chronic obstructive pulmonary disease (COPD), could provide a useful tool to evaluate medical research, particularly if the cited papers contained data on their funding sources—as papers in the WoS routinely do since late 2008.

AMD: Associazione Medici Diabetologi
ANAAO: ASSOMED-Associazione Medici Dirigenti
Consorzio Mario Negri Sud
FAND-AID: Associazione Italiana Diabetici
FIMMG: Federazione Italiana Medici di Famiglia
Gruppo di Studio Complicanze Oculari della Societa Italiana di Diabetologia
SID: Società Italiana di Diabetologia
SIR: Società Italiana della Retina
SOI-APIMO-AMOI: Società Oftalmologica Italiana
Tribunale dei Diritti del Malato

Fig. 22.6 A larger group of Italian nonprofit organizations involved in a diabetes guideline

Fig. 22.7 Sponsors of Spanish CPG in diabetes, including Catalonia

Fig. 22.8 Another example of Spanish diabetes CPG sponsors, including the Basque Country

A major part of the reason for the lack of further use of this tool is undoubtedly that it is difficult to extract the relevant information from the CPGs that would enable each reference to be tabulated with its salient information, such as the authors' addresses and the details of its funding. This is effectively a three-stage process. First, the relevant CPGs have to be found; usually (but not always) they are freely available on the Web. Their titles need to be translated, as those from most non-Anglophone countries will be in the local language. Second, their references have to be identified and collected from the CPG by means of a copying and pasting procedure to a spreadsheet, and then processed so as to give a series of standardized search statements that can be applied to the WoS. Third, they have to be sought in the WoS and their details have to be downloaded to file ready for analysis. This process is described in more detail in Sect. 22.2.1; it is inevitably rather labor-intensive because the different CPGs give references in different formats, and sometimes the references even within one CPG can differ in their format, although a certain degree of assistance can be provided with a visual basic application (VBA) macro. Moreover, some of the references will not be to papers in the serial literature; these cannot be processed in the same way and are usually ignored. A series of VBA programs (written by Philip Roe of Evaluametrics Ltd.) were developed to assist with the extraction of the bibliographic details of the research papers. These VBA macros were specifically formulated to work with the WoS and this was particularly useful as the WoS permits the retrieval of grey literature including conference proceedings not covered in other databases, as well as providing options for accessing address details, country of publication, citation impact metrics, and other parameters in an analyzable format.

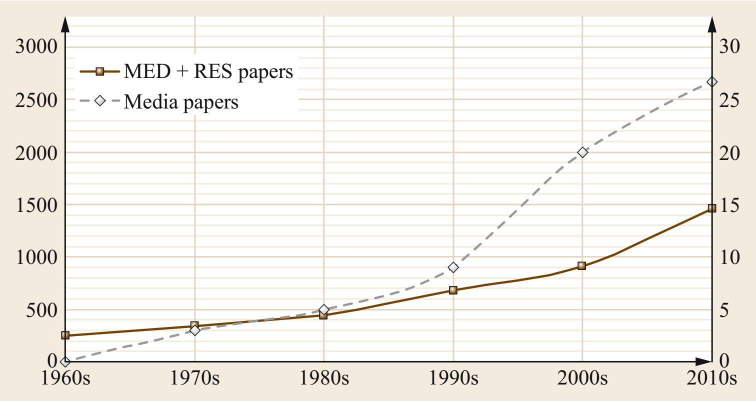


Fig. 22.9 Numbers of papers in the WoS in each decade since 1960 that have medical and research in their titles (*squares* and *solid line*, *left scale*), or concern how medical research is reported in the news media (*diamonds* and *dashed line*, *right scale*)

22.1.4 Newspaper Stories About Medical Research

In contrast to the paucity of papers about the use of the references on CPGs for research evaluation, there are over 50 papers in the WoS that concern how the mass media, particularly newspapers and magazines, report medical research, and their number has been increasing faster than the numbers with *medical* and *research* in their titles (Fig. 22.9).

Of these papers, the majority (31 out of the 52 with addresses) were from the USA, but this may reflect the bias in WoS coverage of social science journals, which accounted for half of the papers. Other countries that contributed papers were continental Europe (8), the UK (7), Canada (4), Japan (2), and China and Russia (1 each). A majority of the papers analyzed stories in newspapers (28) and/or magazines (12); only a handful were concerned with broadcast media, probably because archives of radio and television are less common and less easy to analyze. Many papers looked at how the general public, or sections within it, reacted to mass media coverage of a topic, and especially the effectiveness of public health campaigns [22.35]. Two papers showed that mass media coverage also led to more citations of the research publications as a result [22.36, 37]. The most popular disease area to be analyzed was cancer (25 papers), followed at some distance by mental health, including alcoholism (7), cardiovascular research (2), and arthritis and diabetes (1 each).

22.2 Methodology

A systematic approach was used to identify and download the CPGs, collect the cited evidence-base on these CPGs, download their bibliographical details, and conduct the analysis. Another search strategy was developed to identify the newspaper stories that reported medical research, enter their details to a spreadsheet, and then download the bibliographical details of those research studies for the analysis.

22.2.1 The References on Clinical Practice Guidelines

The first task, of course, was to find the CPGs in the various countries. Since most of them were in languages other than English, we called upon our European partners, other European collaborators, and the KCL graduate students to search for these on the Web, and to provide us with copies, normally in pdf format, with translations of their titles. Altogether we were able to obtain CPGs from 21 countries; some of the others did not appear to have any (e. g., Cyprus at the time, others may have had them, but lack of resources meant that we could not obtain them, e. g., Iceland and Norway). We compiled lists of all these guidelines, and it was immediately apparent that there were far too many for us to be able to process them all, since some had upwards of a thousand references, and many had several hundred, and they tended to be in different formats.

We therefore needed to make a selection, and decided to cover those CPGs that referred to diseases or disorders that were responsible for 1% or more of the disability-adjusted life years (DALYs) in EUR31 in 2010 as given by the Institute for Health Metrics and Evaluation at the University of Washington, USA [22.38]. These were as shown in Table 22.1, and included one or more disease areas from each of the five NCDs.

The next step was to find the references on each of these guidelines. Some CPGs had them neatly gathered

Table 22.1 List of 13 noncommunicable diseases or disorders selected for the analysis of their European clinical practice guidelines, with the estimated percentage European disease burden in DALYs in 2010

Disease area	(%)
Ischaemic heart disease	9.7
Cerebrovascular disease	5.3
Unipolar depressive disorders	4.3
Trachea, bronchus and lung cancers	3.5
Chronic obstructive pulmonary dis.	2.9
Diabetes mellitus	2.5
Colon and rectum cancers	2.0
Anxiety disorders	1.7
Alzheimer's dis. and other dementias	1.7
Breast cancer	1.5
Alcohol use disorders	1.3
Drug use disorders	1.3
Asthma	1.1

together at the end (for example, those from SIGN); others had them at the end of each chapter or section. Naturally, their format varied according to the source of the CPG, and sometimes even within the same document. The normal format included the names of several authors, the title, the year, and the source (journal, volume, pages). However, the Finnish CPGs only gave the name of the first author and did not give the title of the cited paper, and special arrangements were needed for these references (see below).

The reference section was copied and pasted into an Excel spreadsheet. For some CPGs, the references were numbered sequentially, which allowed a specially-written VBA program to identify where each reference ended and the next began (because many ran on to two or more rows in the spreadsheet). For those that did not, the reference section was copied and pasted into MS Word, so that the numbers could be manually inserted before being transferred back to a spreadsheet.

The VBA program then parsed each reference into a standard form of WoS search statement that included up to three authors, plus the three longest words from the reference title, plus the year, plus the initial letter of the journal. (We could not use the journal name because this was often abbreviated in a non-standard format.) An example is given below.

Original Form of Reference

Jensen DM, Damm P, Sorensen B, Molsted-Pedersen L, Westergaard JG, Korsholm L, et al. Proposed diagnostic thresholds for gestational diabetes mellitus according to a 75 g oral glucose tolerance test. Maternal and perinatal outcomes in 3260 Danish women. Diabet Med 2003 Jan;20(1):51–57.

Format when Search Statement Prepared

```
AU=(Jensen, D AND Damm, P AND
Sorensen, B) AND TI=(diagnostic
AND thresholds AND gestational)
AND PY=2003 AND SO=D*
```

These search statements were then grouped automatically into sets of a selected number, typically 20, that could be run against the WoS. References that did not appear to fit the format for journal papers were initially ignored by the program, but could be added later. Sometimes this was because the punctuation was not exact. (For example, the program expected journal references to have just three full stops: one after the list of authors, one after the title, and one at the end. In the example above, there are four because the title consists of two sentences, so this reference would initially have been rejected until the full stop after *test* was removed. Some other references were rejected because the title ended with a question mark.) It was also necessary to check that author names did not have accents, or other diacritical marks such as umlauts or Ø letters, as these are not used in the WoS.

For some sets of references where the numbers had been inserted manually in MS Word, they could each be separated by a paragraph mark, and then when they were transferred to Excel, each was on a single line. This enabled the different elements of the reference to be spread across to different columns, provided that the separators (usually a full stop) could be identified and they were correctly placed in the reference (see above). The `CONCATENATE` function could then be used to prepare search statements using the first author's name, the full title, and the year. The compound search statements were then run against the WoS for all years and all document types. However, sometimes these did not run if syntactic rules were inadvertently broken, such as the inclusion of terms such as *and/or* or the word *near* in the paper title.

Sometimes a compound search statement with, say, 20 individual statements yielded fewer than 20 papers; this was usually because one or more references were not in journals processed for the WoS in the given year. However, the reverse could also occur, with perhaps 22 or 23 papers identified. This was nearly always because the WoS had also recorded corrections or letters to the same journal about the original paper, together with the authors' reply, which would have had the same author(s), title and year, and so satisfied the search statement. We subsequently discarded items described by the WoS as *corrections* and *letters* as they would not have reported research results.

There was a particular problem with the Finnish CPG reference lists. The references were provided as a continuous list, such as this extract:

> 2014;23:39–46 80. Okin PM ym. Hypertension 2000;36:766–73 81. Koren MJ ym. Ann Intern Med 1991;114:45–52 82. Casale PN ym. Ann Intern Med 1986;105:173–8 83. Devereux RB ym. Hypertension 1994;23:802–9 84. Anavekar NS ym. N Engl J Med 2004;351:1285–95

The complete set of references on a CPG was first pasted into a Word document, and then paragraph marks were inserted before each reference number. They were then pasted into Excel, and the string *ym* (the Finnish form of et al.) removed from each where it occurred. The reference was then spread across columns containing the first author, the publication year, the journal name as given, and the volume number and pagination. The journal name had then to be converted into the full journal name with a journal name thesaurus. The resulting search strategies were then run against the WoS, and the papers downloaded to file, up to 500 at a time. These were converted into an Excel spreadsheet by means of another VBA program which put the downloaded data into a standard format for analysis. However, this file contained many additional records by the named author in the given journal and year, and those not conforming to the given page numbers were subsequently removed by hand.

The analysis of the downloaded files was carried out by a further set of VBA programs that carried out the following functions:

- Characterization of each paper as *clinical* or *basic* or *both* according to the presence of one or more words from two lists of selected words in its title [22.39]
- Provision of the fractional counts of each country listed in the addresses
- Identification of the disease area (e. g., cancer site, such as breast or colon, or mental disorder, such as Alzheimer's or depression)
- Identification of the research domain (e. g., genetics, surgery).

This classification enables an understanding of the type of research, whether it is applicable to patients in the clinic, at the laboratory stage, or a mix of both. Research level (RL) is designated by a decimal number between 1.0 = clinical and 4.0 = basic. Each paper cited within a guideline was classified as clinical (1) or basic (4) or both (2.5) and these values were averaged to yield the research level of the set of cited papers, shown as RL (*p*). The same process was repeated to get the average research level value based on the journals in which these papers were cited as described above, shown as RL (*j*)

Information was also available on the gap (in years) between the date of the citing CPG and that of the cited paper [22.33], and on its funding (if it was published in 2009 or later). The latter topic, which was discussed in a recent paper [22.40], will be explored in a later paper.

The references on the clinical practice guidelines for each of the five NCDs were collected together in five separate spreadsheets, and the results are presented separately. Overall comments on them are brought together in the Discussion section. Countries' tendency to cite their own papers on CPGs is examined for all five NCDs.

22.2.2 The Newspaper Stories and the Research That They Reported

Our original intention had been to select one, two, or three newspapers from each of the 31 European

Table 22.2 List of 31 European newspapers whose medical research stories were collected and used to create the file of news stories

Country	Newspaper
Austria	Die Presse
Belgium	De Standaard
Belgium	Le Soir
Bulgaria	Дневник (Dnevnik)
Bulgaria	Труд (Trud)
Croatia	Vecernji List
Cyprus	Cyprus Mail
Czech Republic	Blesk
Denmark	Jyllands Posten
Estonia	Õhtuleht
Estonia	Postimees
Finland	Helsingin Sanomat
France	Le Monde
Germany	Süddeutsche Zeitung
Greece	Το Βήμα (To Bema)
Hungary	Magyar Nemzet
Italy	Corriere della Sera
Italy	La Repubblica
The Netherlands	Het Algemeen Dagblad
The Netherlands	De Telegraaf
Poland	Fakt
Portugal	Correio da Manhã
Portugal	Jornal de Notícias
Romania	Adevarul
Spain	ABC
Spain	El Mundo
Spain	El País
Sweden	Svenska Dagbladet
Switzerland	Berner Zeitung
UK	Daily Mail
UK	The Guardian

Table 22.3 ISO2 codes for countries whose outputs are considered in this study

Code	Country	Code	Country
AT	Austria	IE	Ireland
AU	Australia	IL	Israel
BE	Belgium	IS	Iceland
BG	Bulgaria	IT	Italy
BR	Brazil	JP	Japan
CA	Canada	KR	South Korea
CH	Switzerland	LT	Lithuania
CN	China	LV	Latvia
CZ	Czech Rep.	NL	The Netherlands
DE	Germany	PL	Poland
DK	Denmark	PT	Portugal
EE	Estonia	RO	Romania
ES	Spain	SE	Sweden
FI	Finland	RU	Russia
FR	France	SI	Slovenia
GR	Greece	SK	Slovakia
HR	Croatia	UK	United Kingdom
HU	Hungary	US	United States

Table 22.4 Columns on spreadsheet giving details from selected newspaper stories

Column	Content
A	Index for story and paper
B	Date of newspaper story
C	Country ISO2 code
D	Country name
E	Newspaper code
F	Headline (original)
G	Headline (English)
H	Synopsis (original)
I	Synopsis (English)
J	Length (word count)
K	Journalist name
L	Journalist job title
M	Job sector code
N	NCD code
O	Disease code
P	Research domain code
Q	Researcher(s) named
R	Their institution(s)
S	Journal of cited paper
T	Funding sources
U	Commentator name(s)
V	Commentator institution(s)
W	Notes
X	URL of cited paper
Y	DOI of cited paper
Z	Title of cited paper

countries, with the larger countries being represented by papers with different political outlooks, readership strata, and geographical origins. It was also necessary

Table 22.5 Columns on spreadsheet giving details of cited papers from the Web of Science

Column	Content
AA	Authors
AB	Title
AC	Source
AD	Doc type
AE	Addresses
AF	Publication country
AG	Publication year
AH	Publication month
AI	Language
AJ	Author email(s)
AK	Funders (FU)
AL	Acknowledgement (FX)
AM	Authors and addresses (C1)
AN	Author full names (AF)

for the papers to have an easily searchable archive, or for their full texts to be available on the Factiva © Dow Jones database. However, it proved difficult to find researchers with all the language skills needed (we were eventually able to cover 18 European languages but these did not include Icelandic, Latvian, Lithuanian, Maltese, Norwegian, Slovakian, and Slovene). The time necessary to process newspaper stories for the 12 years of the study period (2002–2013) also prevented us from covering as many newspapers as we would have wished, as it turned out that our search strategies for capturing relevant stories from newspaper archives yielded a large number of false positives, which had to be read individually in order for irrelevant stories to be discarded. (Stories were only retained if they cited research from one or more papers from journals covered in the WoS.) Another limitation was that some newspaper archives did not go back as far as 2002.

Table 22.2 shows a list of the newspapers that were processed. Each newspaper was given a code, consisting of its country ISO2 code (Table 22.3) and one letter, normally the initial letter of its name. Most newspapers had their own searchable databases, but for some the researchers used the Factiva full-text database © Dow Jones. The search strategies used included the names of the relevant NCD diseases or disorders, and a set of terms indicative of research, thus:

```
(cancer OR leukaemi* OR melanoma*
OR lymphoma*) AND (research* OR
study OR scientist* OR expert*)
```

This search strategy, and the four others like it, were all used to search for relevant stories. They were translated into the 17 languages needed to search the non-English language newspapers.

The researchers were all brought to KCL in groups for training. They were asked first to translate the search statements into their own languages, and then to check the selected newspapers' websites for their ability to be searched with the five search statements, and for how many years this could be done. Most of these archives were freely accessible, but for some we needed to subscribe to the newspaper for a short period in order to be granted access. The researchers were also taught about the use of short codes to connote the disease(s) or disorder(s) mentioned in the newspaper stories (three-character, or trigraphs) and the research domain (four-character, or tetragraphs). The NCD was to be connoted by its pentagraph code; some stories covered more than one of them.

The details of each selected newspaper story were then to be copied and pasted, or typed, into a spreadsheet containing 26 columns, as listed in Table 22.4.

The notes item (column W) was simply to assist the researcher to identify the cited paper in the WoS. This was the next task, and each paper that was identified was then downloaded with as identifier the index number of the story. (A few stories cited more than one research paper; these were given consecutive index numbers.) The full records were downloaded, and the details were converted into an Excel spreadsheet by means of the VBA macro used for the references on CPGs: they were then copied across to the spreadsheet of the stories with columns as in Table 22.5.

The cited papers were than analyzed by means of VBA programs in a similar way to that used for the references on CPGs, see the previous section. This added many extra columns to the spreadsheet.

22.3 Results: Clinical Practice Guidelines

In total, we selected 413 CPGs in 26 European countries across the five NCD areas with an evidence-base of 47 274 cited research papers. These were identified from the selected CPGs and their details identified and downloaded from the WoS for analysis.

22.3.1 Clinical Practice Guidelines—Cardiovascular Research and Stroke (CARDI)

We processed 74 CPGs in this disease area from 19 countries; 54 of the CPGs were for coronary heart disease (COR) and 20 were for stroke (STR). (These were the two disease areas that had been selected for study within CARDI, see Table 22.1). There were 11 762 references in total, of which 7447 were for heart disease and 4315 for stroke.

As is usual with the papers referenced on CPGs, the papers were very clinical, with mean RL (p) varying only slightly, from 1.12 for the papers cited by Austrian and Swedish CPGs to 1.03 for those cited by Spanish CPGs. However, the RL of the journals appeared less clinical, and the average RL (j) was 1.36. Some of the journals were more basic: 2110 papers (18%) had RL (j) > 1.5 and 220 of them (1.9%) had RL (j) > 2.5.

Some countries' research is much better cited on the European CPGs than others, and this is shown in Fig. 22.10. This figure shows that the spots for most European countries lie above the diagonal line, and those for East Asian countries lie below it. This is not surprising because the CPGs are all European and there is a tendency for research documents preferentially to cite papers by the countrymen of the document authors [22.33, 41, 42].

Nevertheless, some non-European countries' research is well cited by these CPGs, notably that of Canada, Australia, Israel, and the USA. It might be supposed that this is just because the English-language UK CPGs cite a substantial proportion of the references (2632 out of 11 762, or 22%) but this is not the case, as Fig. 22.11 demonstrates, where these four countries' papers appear to be at least as well cited by Austria + Germany, and by France + Poland.

The tendency of research documents to cite papers by their fellow countrymen is shown by Table 22.6, which shows the over-citation ratio (OCR) for eight European countries and also these countries' presence in CARDI research. The OCR is a measure of the tendency of authors preferentially to cite papers by their fellow countrymen [22.41]. It is greater for countries with smaller scientific outputs, and has also tended to decrease with time as international communication has become easier and cheaper.

Table 22.6 Over-citation ratio for own country papers cited in eight countries' CARDI CPGs

ISO	All	Own	% own	% WoS	OCR
SE	492	56	11.4	2.2	5.24
UK	2632	517	19.6	7.9	2.48
AT	610	18	3.0	1.2	2.39
FR	1437	148	10.3	4.6	2.23
PT	355	3	0.8	0.39	2.17
PL	974	32	3.3	1.6	2.10
ES	433	19	4.4	3.0	1.45
DE	2048	206	10.1	8.5	1.18

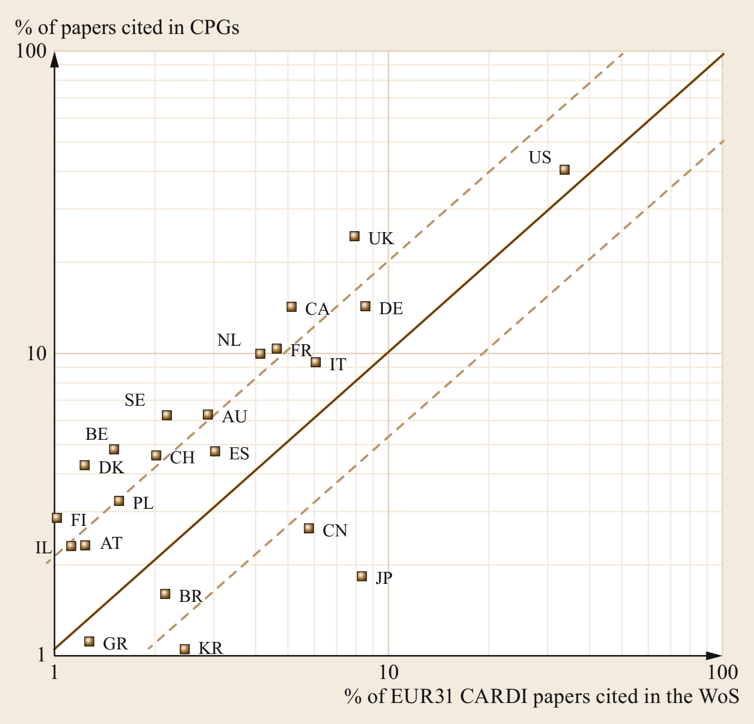


Fig. 22.10 Percentage presence of different countries in papers cited in European clinical practice guidelines for cardiovascular disease and stroke as a function of their percentage presence in CARDI papers in the Web of Science, 2002–2013, integer counts. For country codes, see Table 22.3. *Dashed lines* show percentage presence twice or half that expected on the basis of countries' presence in the WoS

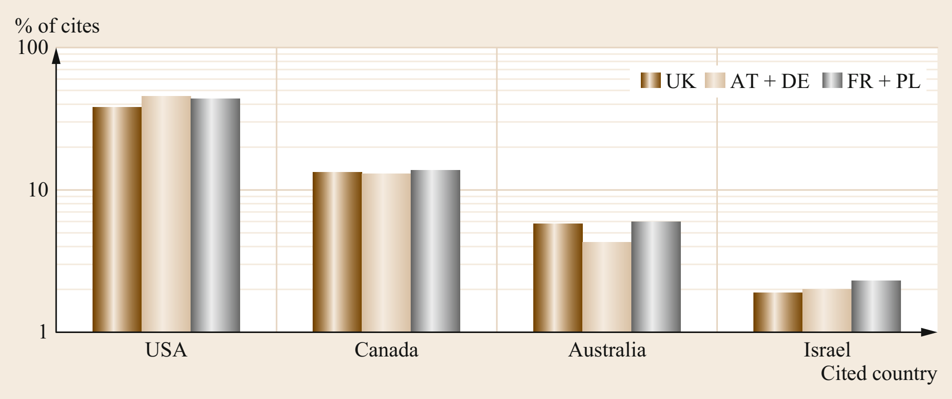


Fig. 22.11 Percentage presence of USA, Canada, Australia, and Israel among the papers cited by CARDI CPGs from the UK ($n = 2632$), from Austria and Germany ($n = 2658$), and from France and Poland ($n = 2411$), integer counts

Some over-citation ratios are somewhat smaller than expected, particularly from Spain and Portugal.

22.3.2 Clinical Practice Guidelines—Diabetes (DIABE)

These guidelines were not divided up by disease area, or by the *sequelae* that often result from diabetes, as the WHO and IHME data on disease burden do not distinguish between them. There was a total of 101 guidelines, see Fig. 22.3, from 25 countries, with a combined total of 5941 references. However, this total included many papers that were cited multiple times on these CPGs, with two papers being cited on as many as 17 of them. Figure 22.12 shows that the distribution of citations follows a logarithmic pattern.

The gap between publication of the CPGs and of the references that they cite is shown in Fig. 22.13 with, for comparison, the distribution by year of synchronous citations (references) on a sample of diabetes papers from 2013, both as percentages of citations over a 19-year

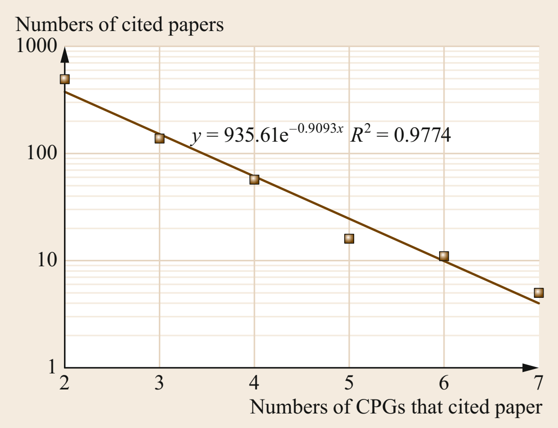


Fig. 22.12 Distribution of citation scores for diabetes papers on 101 European CPGs

period. This suggests that the references on the CPGs are relatively recent: half appeared no more than five years previously. However, the CPG countries varied in how recent their cited references were: this is shown in Fig. 22.14. Finland, Germany, and the UK cite relatively old papers, but Croatia and Portugal relatively recent ones.

The next analysis was of the subject areas of the cited references and a comparison with the subject areas of European diabetes research in 2002–2013. The subject areas were connoted by trigraph codes, listed in Table 22.7 and the comparison is in Fig. 22.15.

Table 22.7 List of diabetes research subject areas, with trigraph codes

Code	Subject area	Code	Complications
TY1	Type 1	FEE	Feet
TY2	Type 2	CAR	Cardiovascular
GES	Gestational diabetes	KID	Nephropathy
NEO	Neonatal diabetes	NEU	Neuropathy
MOD	Maturity onset diabetes of the young	LIV	Liver
ADA	Latent autoimmune diabetes of adults	HYP	Hypoglycaemia
RET	Complications: retinopathy	PSY	Psychosocial
		GEN	Genetics

It appears that there is a reasonable match between the subject areas in which European diabetes research is being undertaken and those that are important in the provision of the evidence base. This is not the case for cancer, as we shall see. The subject areas that are of less utility in the provision of this evidence are genetics and effects on the liver. The latter may be due to there being few CPGs covering this subject area. Conversely, there is a lack of research on the effects of diabetes on the feet, which appears important for clinical practice.

The tendency for countries to cite their own papers among the references on their clinical practice guidelines is examined in Table 22.8. The ratios are higher than they were for the CARDI CPGs, and Spain (with Portugal) and Germany are again citing their own papers less often than do the other countries.

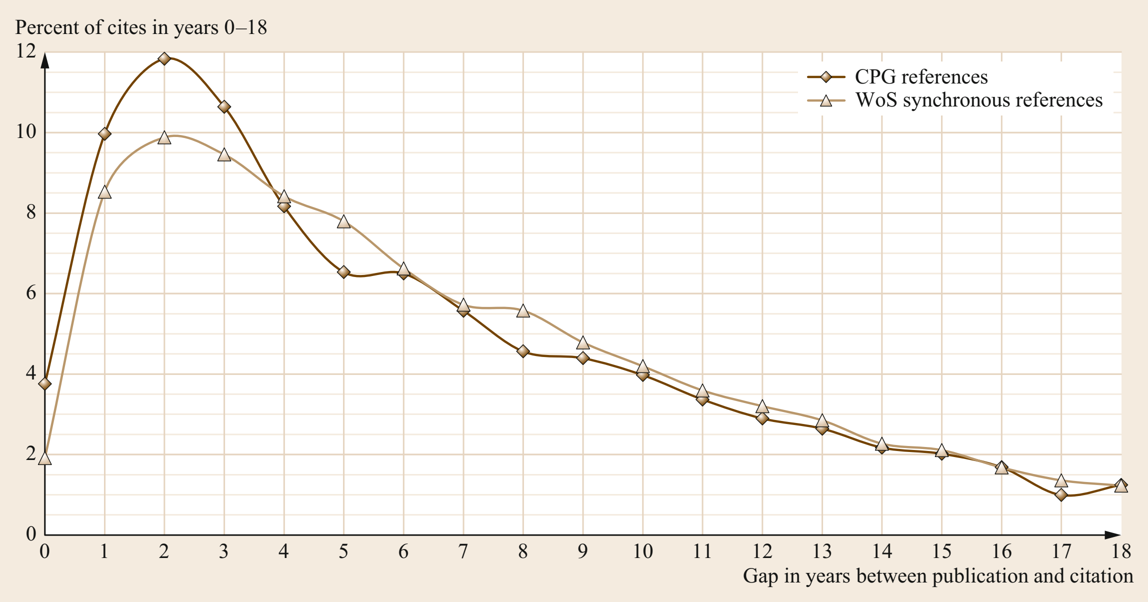


Fig. 22.13 Gap between dates of cited references on diabetes clinical practice guidelines and the guidelines (*light brown line*), and comparison with the time distribution of synchronous citations in 2013 from diabetes papers in the WoS (*dark brown line*)

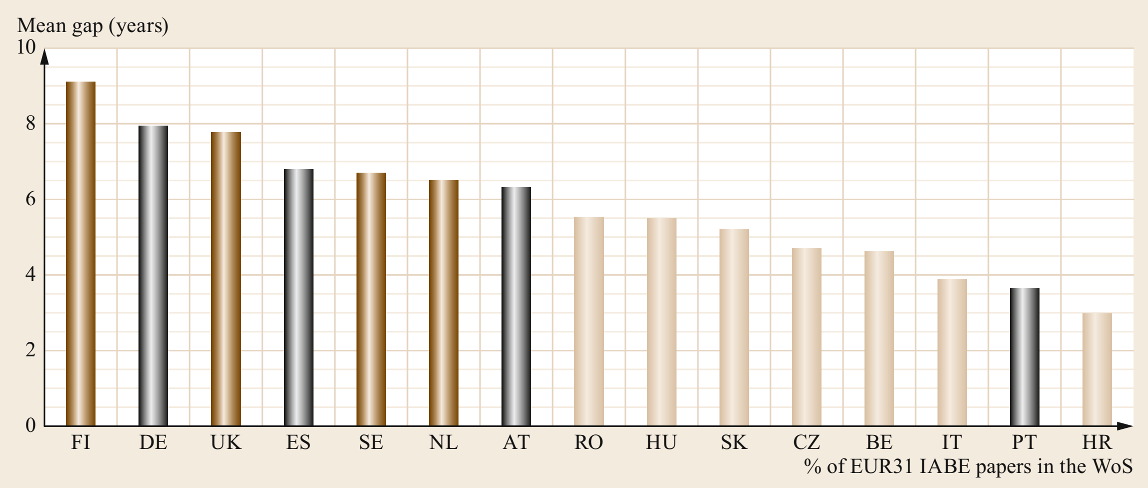


Fig. 22.14 Mean gap between publication of a clinical guideline for diabetes and the references that it cited, for 15 European countries whose guidelines cited at least 75 references. Countries with > 600 cited references shown by *brown bars*; countries with > 250 cited references shown by *gray bars*; with < 250 references shown by *beige bars*

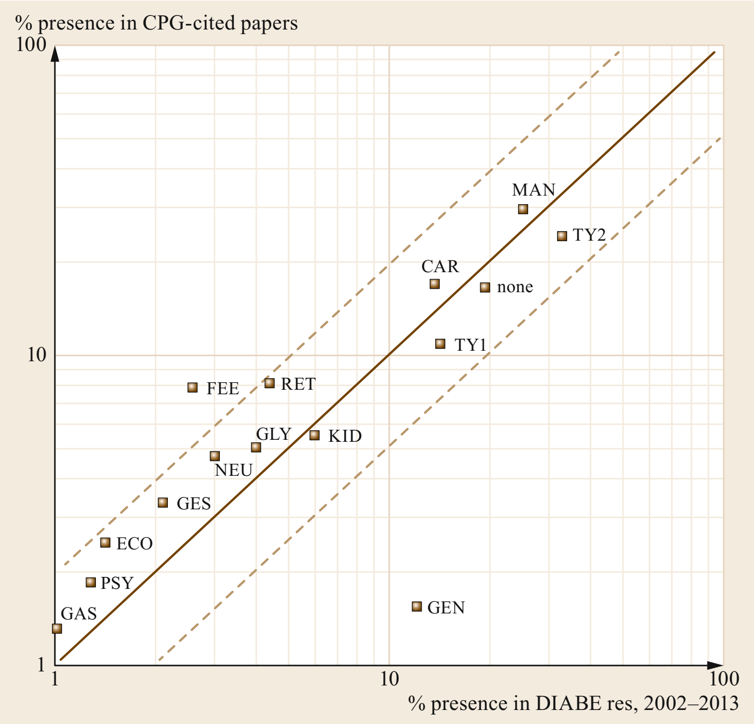


Fig. 22.15 Relationship between European diabetes research subjects, 2002–2013, and the evidence base of 101 European CPGs. For subject area codes, see Table 22.7

22.3.3 Clinical Practice Guidelines—Mental Disorders (MENTH)

This NCD differs from the others because the individual disorders are so different from each other. The CPGs selected for analysis covered five disorders: drug use disorders (addiction) (ADD); alcohol misuse disorders (ALC); Alzheimer's disease and other dementias (ALZ); anxiety disorders (ANX); and unipolar depressive disorders (DEP). Altogether, we analyzed 100 CPGs from 20 European countries citing 12 442 research papers and their division between the disorders, together with the numbers of references, the European disease burden, and the amount of European research,

Table 22.8 Over-citation ratio for own country papers cited in ten countries' DIABE CPGs

ISO	All	Own	% own	% WoS	OCR
AT	406	46	11.3	1.1	10.2
FI	633	86	13.6	1.7	7.91
NL	646	126	19.5	3.1	6.23
BE	143	6	4.2	1.2	3.48
SE	340	38	11.2	3.3	3.41
UK	948	271	28.6	9.2	3.10
IT	177	29	16.4	5.4	3.04
DE	562	96	17.1	6.6	2.59
ES	1175	53	4.5	2.9	1.53
PT	262	1	0.4	0.4	1.04

is shown as a chart in Fig. 22.16. There appears to be a reasonable correspondence between the four parameters for each disorder: depression attracts the most research activity, and there are more references per CPG than for the other disorders.

The presence of the leading countries among the cited references on these CPGs is compared with their presence in mental disorders research in Table 22.9.

The UK, Sweden, Finland, and the Netherlands show to advantage here, and the non-European countries in Asia and South America are less cited, including Israel. Country self-citation ratios are rather higher than they were for CARDI (Table 22.10). It is striking that Germany and Spain cite their own papers less than do the other countries, as was the case for the CARDI and DIABE CPGs.

22.3.4 Clinical Practice Guidelines—Cancer (ONCOL)

The three most serious cancers—lung, breast, and colorectal—were the ones for which CPGs were selected. There were a total of 81 ONCOL CPGs: 31 for breast cancer (MAM, 3748 references), 30 for lung cancer (LUN, 4319 references), and 20 for colorectal cancer (COL, 1773 references). Figure 22.17 shows a log–log plot of 18 countries' presence among the cited references compared with their presence in cancer research in the WoS for 2002–2013.

The same pattern appears as with CARDI papers: most European countries' papers are relatively over-cited, and those from the three East Asian countries (China, Japan, Korea) are under-cited, here by a factor of about two. Belgian papers are the most cited relative to their presence in the WoS, followed by those of the Netherlands, Canada, and the UK. Each country's

Table 22.9 Percentages of countries' papers in mental disorders research, 1995–2011, and among the cited references on European CPGs for mental disorders (CPG cites), and ratio between them. Country codes in Table 22.3. Cells with values: > 2.0 *shaded bright green*; > 1.41 *shaded pale green*; < 0.71 *shaded yellow*; < 0.5 *shaded pink*

Country	WoS papers	CPG cites	Ratio
US	43.3	53.1	1.23
UK	10.0	20.6	2.06
DE	8.5	6.0	0.71
CA	5.6	7.5	1.34
JP	4.6	1.5	0.33
AU	4.3	5.2	1.21
IT	3.7	4.5	1.21
FR	3.7	4.0	1.07
NL	3.3	5.2	1.57
ES	2.9	2.5	0.87
SE	2.4	4.4	1.80
CN	2.2	0.5	0.21
CH	2.0	2.2	1.15
BR	1.6	0.5	0.33
IL	1.3	0.9	0.68
FI	1.2	2.1	1.70
BE	1.2	1.6	1.28
KR	1.1	0.2	0.22

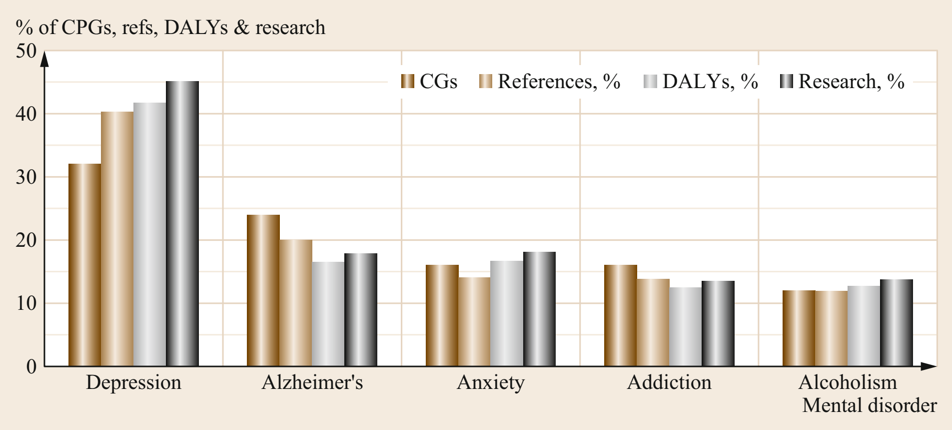


Fig. 22.16 Numbers of European mental disorders (MENTH) clinical practice guidelines, and the percentages of references, burden in DALYs, and research outputs for each of five disorders

Table 22.10 Over-citation ratio for own country papers cited in ten countries' MENTH CPGs

ISO	All	Own	% own	% WoS	OCR
DK	268	24	9.0	1.1	8.46
FI	871	70	8.0	1.2	6.99
SE	1101	133	12.1	2.3	5.22
FR	265	30	11.3	3.5	3.24
UK	3289	931	28.3	10.9	2.59
BE	294	9	3.1	1.3	2.31
NL	1068	93	8.7	4.0	2.20
LT	1150	1	0.1	0.1	1.61
ES	2029	98	4.8	3.3	1.48
DE	1125	87	7.7	8.3	0.93

CPGs over-cite their own countrymen's papers by factors shown in Table 22.11.

Although Germany has a low OCR value, as it does in the other NCDs, Spain and Portugal have much higher ones than usual, suggesting that their cancer research is of greater utility than their research on the other NCDs. Finland and Sweden are again quite reliant on their own research.

The next analysis was of the research domains of the papers that were cited on the CPGs, compared with those of European cancer research in 2002–2013. This is shown in Fig. 22.18, where the abscissa is the

Table 22.11 Over-citation ratio for own country papers cited in 12 countries' ONCOL CPGs

ISO	All	Own	% own	% WoS	OCR
FI	226	26	11.5	0.9	12.9
ES	474	95	20.0	2.4	8.34
PT	264	6	2.3	0.3	7.86
PL	131	7	5.3	0.9	5.69
NL	1665	259	15.6	2.9	5.36
SE	555	43	7.7	2.0	3.86
UK	2481	570	23.0	6.9	3.33
BE	637	28	4.4	1.3	3.28
FR	570	77	13.5	5.2	2.61
IT	1195	176	14.7	5.8	2.54
DE	1068	164	15.4	8.0	1.92
LV	427	0	0.0	0.0	0.00

percentage presence of each of 11 research domains and the ordinate is their presence in the 9840 CPG references.

The main conclusions are two-fold. Genetics research, which is by far the most popular research domain, is of little importance to the development of most CPGs. Conversely, surgery and radiotherapy, which are the main means of curing cancer (as opposed to its palliation), are of great importance for CPGs but are less popular with researchers, are not well funded, and are

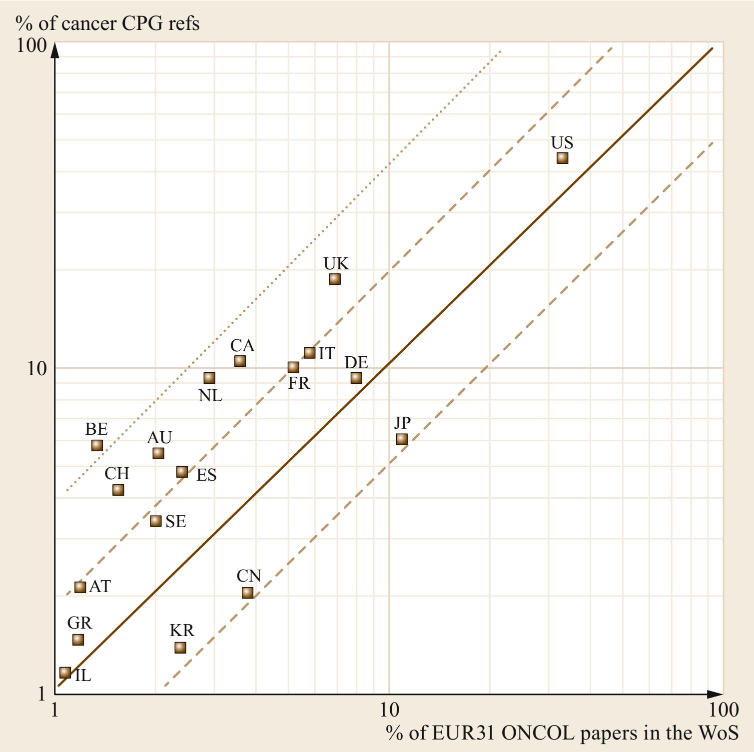


Fig. 22.17 Percentage presence of different countries in papers cited in European clinical practice guidelines for cancer as a function of their percentage presence in ONCOL papers in the Web of Science, 2002–2013, integer counts. For country codes, see Table 22.3. *Dashed lines* show percentage presence four times, twice or half that expected on the basis of countries' presence in the WoS

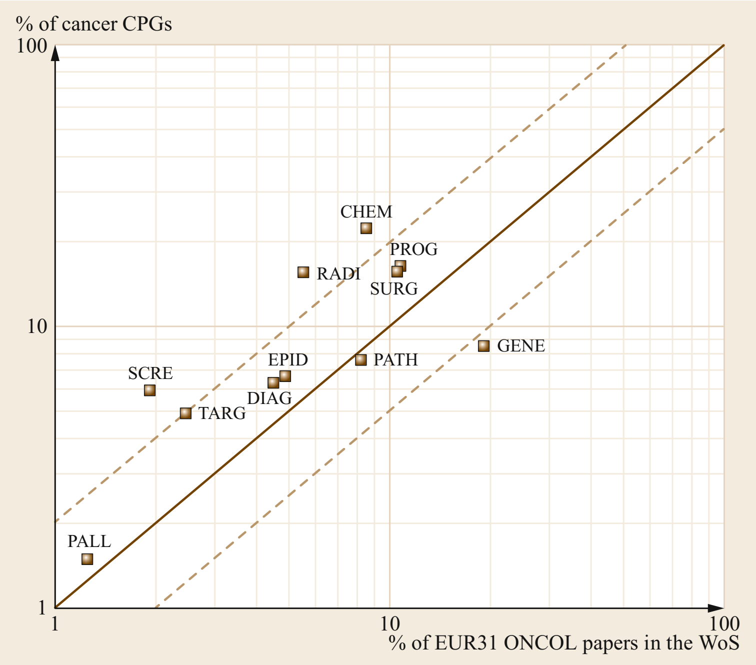


Fig. 22.18 Comparison of presence of each of 11 cancer research domains in EUR31 papers, 2002–2013, and corresponding percentages in the references on European ONCOL CPGs. *Dashed lines* show values twice and half the expected values. Domains: CHEM = chemotherapy, DIAG = diagnosis, EPID = epidemiology, GENE = genetics, PALL = palliative care, PATH = pathology, PROG = prognosis, RADI = radiotherapy, SCRE = screening, SURG = surgery, TARG = targeted therapy

Table 22.12 Comparison between countries' presence in RESPI research in the WoS and their presence in the references in European RESPI CPGs. Values in cells > 2.0 *shaded bright green*; if > 1.41 *shaded pale green*; if < 0.71 *shaded pale yellow*; if < 0.5 *shaded pink*

ISO2	% WoS	CPGrefs	% CPG	OCR
US	34.8	2493	35.1	1.01
UK	21.3	1878	26.5	1.24
CA	7.4	1060	14.9	2.02
FR	6.2	315	4.4	0.71
DE	6.1	388	5.5	0.89
IT	5.7	463	6.5	1.15
AU	5.3	432	6.1	1.15
NL	4.9	652	9.2	1.89
JP	3.9	121	1.7	0.43
ES	3.7	485	6.8	1.83
SE	3.3	388	5.5	1.65
CN	3.2	79	1.1	0.35
BR	2.4	91	1.3	0.53
BE	2.3	296	4.2	1.81
KR	2.2	16	0.2	0.10
TR	2.0	28	0.4	0.20
DK	1.9	335	4.7	2.44
CH	1.7	188	2.6	1.56
PL	1.3	99	1.4	1.04
FI	1.2	159	2.2	1.84
GR	1.2	58	0.8	0.69
NZ	1.1	168	2.4	2.09
TW	1.1	12	0.2	0.15
NO	1.0	149	2.1	2.02

Table 22.13 Over-citation ratio for own country papers cited in nine countries' RESPI CPGs

ISO	All	Own	% own	% WoS	OCR
CZ	178	6	3.4	0.4	9.50
CH	108	13	12.0	1.7	7.10
FI	551	45	8.2	1.2	6.69
ES	1144	179	15.6	3.7	4.18
NL	483	64	13.3	4.9	2.72
SE	143	11	7.7	3.3	2.32
UK	1162	357	30.7	21.3	1.44
FR	213	14	6.6	6.2	1.06
DE	820	51	6.2	6.1	1.02

not well cited in the literature [22.33, 43, 44]. Screening is also under-researched compared with its contribution to the evidence base of cancer CPGs.

22.3.5 Clinical Practice Guidelines—Respiratory Diseases (RESPI)

This was by far the smallest of the five NCDs that we investigated [22.45]. Most of the research was on just two diseases: asthma and chronic obstructive pulmonary disease (COPD). Asthma was covered by 27 CPGs with 3334 references; COPD by 30 CPGs (two covered both diseases) and 4014 references. One CPG covered pulmonary fibrosis, with 94 references. Altogether, 57 CPGs from 19 countries were processed, with a total of 7289 references.

The comparison between research output from the different countries and their presence on the cited references is shown in Table 22.12.

Once again, the Asian countries' outputs and ratios to their presence among the references on RESPI CPGs are all < 0.5 (and are shaded pink), and eight European countries have values > 1.41 (and are shaded bright or pale green).

The over-citation ratios for the countries whose CPGs have at least 100 references and at least one from their own country are shown in Table 22.13. The values for the larger countries are lower than for the other four NCDs, probably because European RESPI research has a bigger presence in the world (56%) than its output in the other NCDs, which averaged 40%.

22.4 Results: Newspaper Stories

From the 31 newspapers covered across 22 European countries, there were 8596 cited research studies featured in newspaper stories for the five NCDs. The details of the stories and of the cited research papers were entered into a single spreadsheet for analysis.

22.4.1 The Five Noncommunicable Diseases

The file of stories and cited papers contained 8596 entries, so it was comparable in size with the sets of references on each of the CPGs. Of these, 3498, or nearly 41%, concerned cancer, and the analysis of these stories and papers is treated in detail in the next section. Here we examine the distribution of the stories between the five NCDs, how it compares with the amount of research on each of them, and their relative disease burden in Europe. We also consider the geographical distribution of the cited papers and their research level.

The numbers of stories were very unequal between countries: the UK and Belgium, each with two newspapers, were much the most productive of data, and there were very few from Austria, Switzerland, and Cyprus (Fig. 22.19). It is not clear if these differences really reflect the amount of interest in medical research in the different countries, or are an artefact of the selection process.

As mentioned above, of the five NCDs cancer was the disease area most often mentioned. Figure 22.20 shows that this reflects the amount of research into the disease in the EUR31 countries, but exaggerates its burden. This is also the case for diabetes, although the burden from this NCD is increasing. Conversely, cardiovascular disease research is under-reported, as is respiratory disease, though the latter is clearly seriously under-researched [22.45]. Mental disorders appear to be getting a fair share of news space.

The research level of the papers cited by the newspaper stories was fairly similar to that of the European research in 2002–2013, except in CARDI and ONCOL where it is somewhat more clinical. This means that the newspapers are selecting research from the whole spectrum of RL for their stories, in contrast to the papers cited as references on CPGs which are almost entirely clinical (Fig. 22.21).

For all the newspaper stories, the over-citation ratios for the leading countries are shown in Table 22.14.

One of the features of the newspaper stories is that many of them included a comment on the significance of the results from an external expert. Altogether,

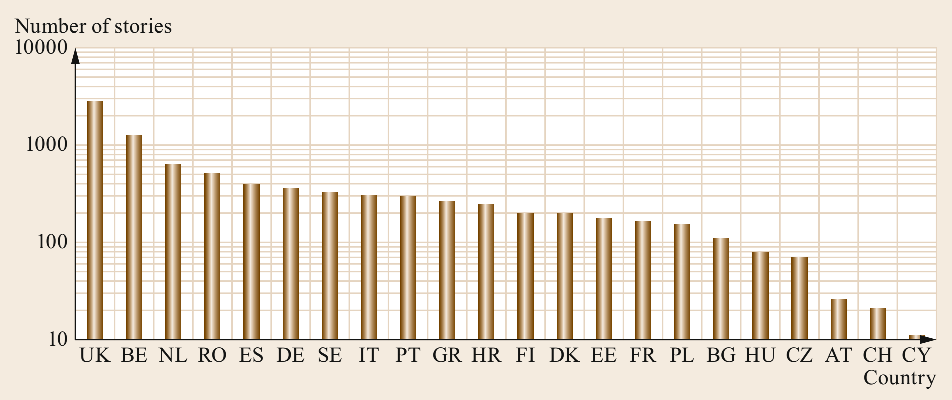


Fig. 22.19 Numbers of newspaper stories about NCD research for 22 countries in 2002–2013. For country ISO codes, see Table 22.3

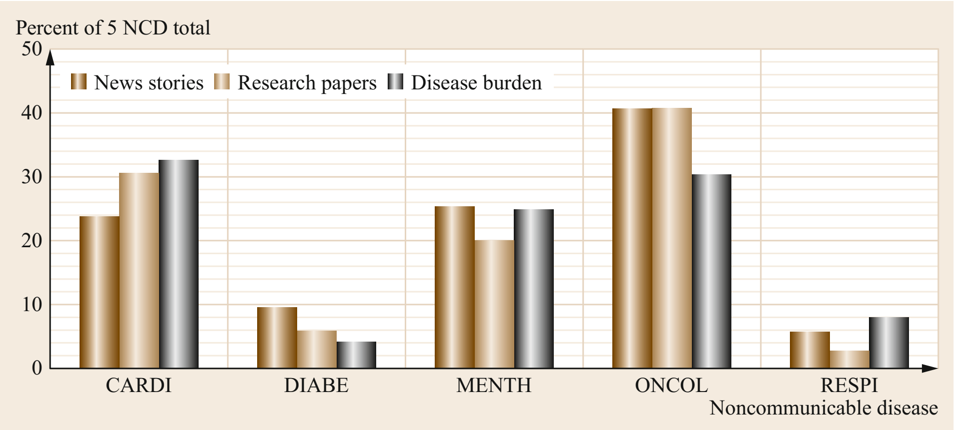


Fig. 22.20 Comparative coverage by newspapers of research on five NCDs compared with the volume of European research (2002–2013) and relative disease burden (DALYs, 2012)

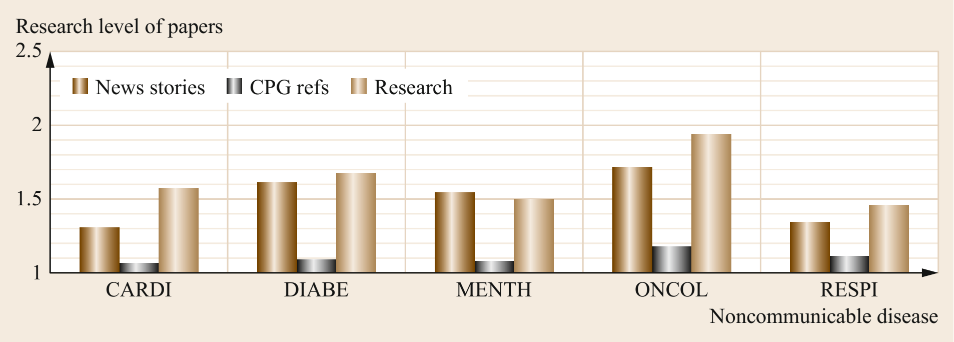


Fig. 22.21 Mean research levels of papers in five NCDs: cited by news stories, cited by CPGs, and published by EUR31 researchers. RL = 1.0 is clinical observation; RL = 4.0 is basic research

Table 22.14 Over-citation ratio for 13 sets of own country papers cited in newspaper stories

ISO	Own	% own	% WoS	OCR
PT	65	21.7	1.2	18
FI	101	50.8	5.1	10
HR	6	2.4	0.3	7.1
RO	13	2.5	0.4	6.1
DK	67	33.8	5.8	5.8
ES	128	32.2	6.9	4.7
NL	298	47.2	10.3	4.6
SE	139	42.6	9.3	4.6
IT	118	38.8	8.9	4.4
GR	24	9.1	2.3	3.9
BE	179	14.3	4.7	3.0
DE	80	22.3	9.9	2.3
UK	1187	42.5	29.7	1.4

1520 stories mentioned a commentator (18%), but the percentages were much higher in the UK (900 with commentators, 32%), Denmark (60, 30%) and Sweden (92, 28%). There did not appear to be any in the German or Italian newspaper stories, but this may have been because the researchers simply did not record them. Table 22.15 lists the ones that were mentioned most often: the list is dominated by UK medical research charities, who are frequently invited to comment by the journalists on the *Daily Mail* and *The Guardian*. However, in other countries, most of the commentators are academics, many of them from the USA.

22.4.2 Mental Disorders Research Stories and Their Cited Papers

Within the subject area of mental disorders (MENTH), where there were 2175 stories and cited papers, the disorders of greatest interest to the journalists were Alzheimer's and other dementias, and depression. This accords with the volume of research, see Fig. 22.22, but it is not in accord with the burden (in DALYs in 2012).

The figure also reveals that alcohol misuse is a more serious problem than all the other disorders listed in the figure, but is relatively neglected both by researchers [22.46] and by the newspapers. There

Table 22.15 List of commenting organizations in newspaper stories about NCD research

Commenting organization	*N*
Cancer Research UK	196
British Heart Foundation	107
Alzheimer's Society (UK)	71
Diabetes UK	41
Alzheimer's Research Trust (UK)	34
Karolinska Institutet (SE)	32
Breakthrough Breast Cancer (UK)	26
Prostate Cancer UK	21
Asthma UK	19
Stroke Association (UK)	19
UK Department of Health	16
Breast Cancer Campaign (UK)	14
University of Leuven (BE)	14
National Public Health Institute (FI)	12
Breast Cancer Care (UK)	11
Erasme Hospital Brussels	11
International Agency Research on Cancer	11
UK Medical Research Council	11
National Health Service (UK)	11
GlaxoSmithKline plc	9
Medicines and Healthcare Products Regulatory Agency (UK)	9
University of Lund (SE)	9
British Thoracic Society	8
Harvard Medical School	8
Institute of Cancer Research (UK)	8
King's College London	8
University of Louvain (BE)	8

Table 22.16 Countries authoring papers cited in newspaper stories about mental health (newspaper stories (NS) cites) and percentage of MENTH papers from each in the WoS, 2002–2013. For country codes, see Table 22.3. Cells with ratio > 2.0 *shaded bright green*; for ratio > 1.41 *shaded pale green*; for ratio < 0.71 *shaded pale yellow*; for ratio < 0.5 *shaded pink*

ISO2	% WoS	NS cites	% NS	Ratio
US	41.9	1152	53.0	1.26
UK	10.6	578	26.6	2.50
DE	8.0	175	8.0	1.00
CA	6.2	180	8.3	1.33
AU	5.2	96	4.4	0.85
IT	4.1	102	4.7	1.16
JP	3.9	44	2.0	0.52
NL	3.8	183	8.4	2.22
FR	3.6	115	5.3	1.46
CN	3.5	36	1.7	0.48
ES	3.4	92	4.2	1.26
SE	2.3	166	7.6	3.27
BR	2.1	12	0.6	0.27
CH	2.0	77	3.5	1.75
KR	1.6	13	0.6	0.38
IL	1.4	21	1.0	0.67
TR	1.4	4	0.2	0.13
BE	1.4	62	2.9	2.04
IN	1.4	13	0.6	0.44
TW	1.3	6	0.3	0.22
FI	1.1	115	5.3	4.73
NO	1.1	67	3.1	2.79
DK	1.1	58	2.7	2.51
PL	1.0	13	0.6	0.63

is also a lack of attention to suicide and self-harm, which accounts for almost 10% of all mental health problems.

The countries authoring the papers cited in the news stories about mental disorders were, as expected, mostly from European countries and the USA (Table 22.16). The comparator is the countries' output of MENTH papers in 2002–2013.

It appears that the best-cited countries are the ones in Scandinavia, together with the UK, the Netherlands, and Belgium. (The value for Iceland is 11.6).

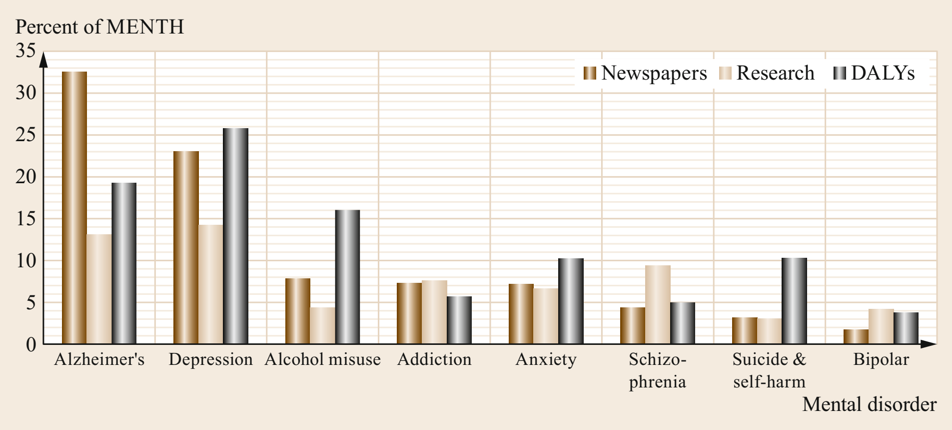


Fig. 22.22 Percentages of news stories, of EUR31 research, and of the disorder burden within MENTH

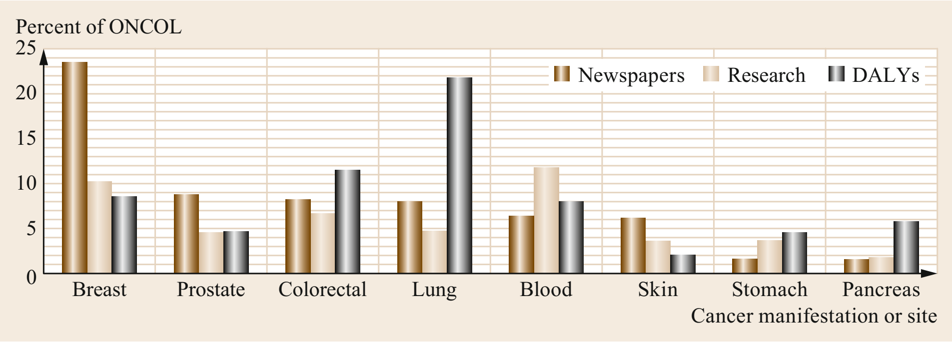


Fig. 22.23 Percentages of cancer newspaper stories, of ONCOL research, and of cancer DALYs, from eight leading body sites

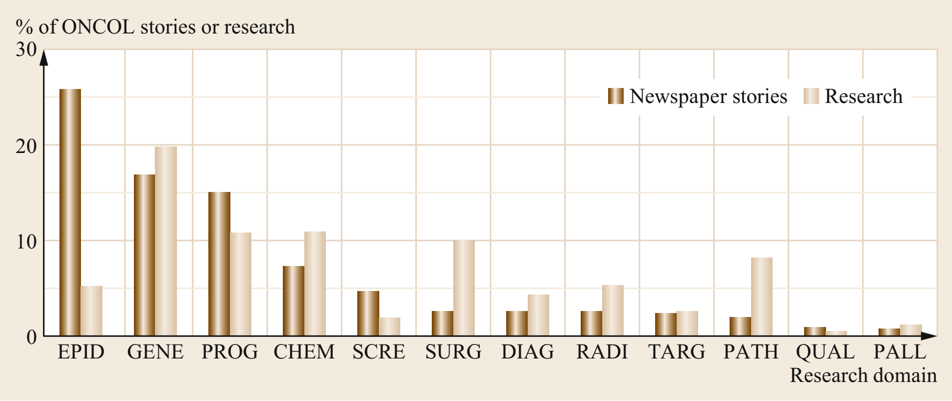


Fig. 22.24 Percentages of cancer newspaper stories and of ONCOL research in the EUR31 countries, on 12 research domains. EPID = epidemiology, PROG = prognosis, GENE = genetics, QUAL = quality of life, PALL = palliative care, SCRE = screening, CHEM = chemotherapy, SURG = surgery, DIAG = diagnosis, PATH = pathology, RADI = radiotherapy, TARG = targeted therapy

22.4.3 Cancer Research Stories and Their Cited Papers

We carried out a separate analysis of the cancer sites most often mentioned in the stories about cancer research, and also of the research domains. Figure 22.23 shows the leading sites mentioned in the stories, with, for comparison, the relative percentages of EUR31 cancer research and of the disease burden in Europe in 2012 in DALYs. There is clearly an imbalance in the selection of stories: breast and skin cancer (melanoma) get more coverage in the newspapers than they merit, but lung and pancreatic cancers get little coverage.

Figure 22.24 shows a similar comparison between the different research domains (here, based on the cited papers rather than the codes given by the researchers) and the amount of research by the EUR31 countries. Coverage is again unbalanced: the main topic of the stories is epidemiology and some distance behind come genetics and prognosis.

However, there is little coverage by the news stories of the three main methods of treatment—chemotherapy, and especially surgery and radiotherapy—suggesting that the main interest of the journalists is in the prevention of cancer rather than its treatment. Of the treatment methods, chemotherapy and targeted therapy combined are covered in 230 stories, but radiotherapy and surgery combined in only 166. This may well give the public the false impression that cancer is cured by drugs rather than the latter treatments [22.43, 44]. To their credit, the journalists do provide better coverage of screening than the researchers do.

22.5 Discussion

In this research work, we evaluated two indicators of research, specifically citations in the clinical practice guidelines and in newspaper stories. This is the first pan-European study to our knowledge that aimed to inform the research gap between research publications and impact on clinical practice through CPGs or public health awareness through newspapers. Furthermore, through a systematic search methodology, this study covered five noncommunicable disease areas, which again, make it perhaps the largest of its kind through an assessment of two research indicators (clinical practice guidelines, newspapers), multicountry comparison, and different disease and subdisease areas pattern examination. Further research on how best the research evidence base or reporting in newspapers can correspond to the disease pattern affecting each European country can perhaps influence clinical practice as well as inform more effective health-policy practices.

22.5.1 Limitations of This Study

The first limitation of this study concerns the inevitable selection of sources, of both CPGs and newspapers. This was constrained by the time and the resources available for the study. We were not unduly constrained by language, as King's College London has graduate students from nearly all continental European countries, and we were able to employ them for the short periods needed for their assignments. However, the training provided to the researchers who were responsible for identifying and processing the newspaper stories, and the papers that they cited, was inevitably rather brief. Some of them may not have fully understood all the complexities of the coding system, or the need for a cited paper to be found in the WoS for its citing story to be included, or indeed for the details of commentators to be recorded. We were able to clean some of the data and this task enabled the results from a few countries to be much better as a result. Some of the countries' newspaper stories did seem rather few in number but much of the analysis presented here is based on results from those countries that appeared to have good coverage of medical research, notably the UK and Belgium.

Many of the results presented were compared with other outputs, notably the amount of research carried out on the five NCDs in Europe, and in other countries. We had to select a time frame for these outputs. For most of the comparisons, we used the 12-year study period (2002–2013). This is probably fair for the newspaper stories, as nearly all of these are written about new research that has just been published, but is more problematic for the papers cited on the CPGs, where some references go back many years. There is no right answer for the appropriate time frame to be used, and the use of the last 12 years meant that the outputs of east Asian countries, most of which have increased rapidly (but not those of Japan), may have put them at a comparative disadvantage.

We found a similar difficulty with regard to the disease burden. There are two main sources of data: the World Health Organization (WHO), and the Institute for Health Metrics and Evaluation at the University of Washington. Both have changed their data from time to time as a result of (presumably) better methods of analysis. In particular, there is inevitably dispute about the weights that should be assigned to disabilities that provide the basis of DALYs. A particular disability may prove much more burdensome in some countries than in others, and the methods used make international comparisons difficult. For instance, the mental disorder DALYs depend critically on the severity of the particular condition (e. g., depression), and this is quite hard to determine.

22.5.2 Advantages of This Study

Despite the above reservations, this study is the first to have compared the references on CPGs and in newspaper stories on five major disease areas and in over 20 countries, with many different languages. It was brought about by the award of a contract by the European Commission, and led to a multinational research activity to find out about the outputs and impacts of European medical research using a standard methodology in these different countries. We are now bringing out a series of papers [22.40, 42, 45] on the individual disease areas and on the methods of analysis that we have developed that will for the first time show the strengths and weaknesses of European medical research, which is inevitably very fragmented and needs to be better coordinated if it is to be efficient. We have also developed a methodology for the measurement of impacts in the real world that can be compared with the traditional evaluation criteria of citation counts [22.47]. These appear rather one-dimensional in comparison and because they are used so much for the allocation of research grants may distort research priorities.

The methodology described here can, in principle, be used on a much wider scale to provide research funders and research performers with information on how their outputs have influenced medical care through CPGs and the public through the mass media. We are considering how best to develop these information sources commercially. The difficulty is that the varied nature of the source materials makes it hard to automate the process of collection of reliable information, and therefore the cost of data collection would need to be spread across many potential subscribers.

22.5.3 Main Conclusions of the Study

Perhaps the most important conclusion is that the papers cited on CPGs and in newspaper stories are not the same ones as receive many citations in the serial literature, and that some relatively neglected areas are unexpectedly important for the practice of medicine or for the provision of useful information to the public. In cancer, surgery and radiotherapy have emerged as important areas for the guidance of physicians and surgeons.

In mental health, the public perception of depression as a subject that used to be kept under wraps has changed, and there is now a willingness to accept it as an illness that can and should be treated. Similarly, the treatment of Alzheimer's and other dementias is getting increasingly more attention, and this is aided (in the UK) by the prominence of the two Alzheimer's medical research charities as commentators on mental health stories. This will help them in their mission to raise funds to support new research.

The solicitation of comments from collecting charities by the UK media, which was noted earlier [22.37], occurs in other disease areas, as witness the prominence of Cancer Research UK, the British Heart Foundation, and Diabetes UK among the leading commentators (Table 22.15). It would surely be helpful to the corresponding charities in other European countries if journalists could call on them regularly (and expeditiously) to comment on the news stories that they were planning to write. This seems to be happening on a small scale in Denmark and the Netherlands, but very little elsewhere.

We also noted the over-citation of the research papers by their authors' fellow countrymen. Some countries, particularly in Scandinavia, had high observed-to-expected ratios of citations. This was associated with a small percentage presence in the world literature of a subject area.

Acknowledgments. This study was supported by the European Commission through the award of a contract to the London School of Economics and Political Science (EC/FP7/602536). It was very materially assisted by the provision of several VBA programs for use with MS Excel for the collection and analysis of data; these were written by Philip Roe of Evaluametrics Ltd.

The CPGs and newspaper stories collection, and identification of the cited research studies from the newspapers, was done by the following individuals: for Austria, Germany, and Switzerland: Natalia Kelsch, Anne Spranger, Victor Stephani, and Tobias Schumacher from Technische Universität Berlin, Germany; for Belgium: Ann-Sophie de Mol and Gabrielle Emanuel from King's College London (KCL), UK; for Bulgaria: Eva Nacheva and Christina Tencheva from KCL; for Croatia: Ria Ivandic Emanuel from KCL; for Cyprus: Chryso T. Pallari from the University of Cyprus, Nicosia, Cyprus; for the Czech Republic and Poland: Kasia Zemanek from KCL; for Denmark: Maria Dahl and Maria Emilsson from KCL; for Estonia: Argo Soon from the Estonian Research Council, Tartu, Estonia; for Greece: Laura Mantovani from KCL; for Hungary: Csajbok Edit from Semmelweis University, Budapest, Hungary; for Italy: Ludovica Borsoi from Università Commerciale Luigi Bocconi, Milan, Italy; for Latvia: Ingrid Jaselskyte, Estonian Research Council, Tartu, Estonia; for Luxembourg and the Netherlands: Ann-Sophie de Mol from KCL; for Portugal: Diana Gosálvez-Prados, Elisabeth María Ilidio-Paulo, Camila Higueras-Callejón, and José Carlos Ruiz-Jiménez from Escuela Andaluza de Salud Publica, Granada, Spain; for Romania: Maria-Cristina Juverdeanu from KCL; for Spain: Diana Gosálvez-Prados and Elena Salamanca-Fernández from Escuela Andaluza de Salud Publica, Granada, Spain and Tahereh Dehdarirad from Universitat de Barcelona, Barcelona, Spain; for Sweden: Gustaf Nelhans from the University of Boras, Sweden; for the UK: Argo Soon, Marleen Saidla and Tiina Tasa (Estonia) and Eva Nacheva from KCL partly assisted EP on the data collection.

22.A Appendix

Table 22.17 Organizations that publish CPGs in Europe

Country	Coordination	Organization(s)	Website
Austria	Decentralized with various organizations involved	Scientific associations of medical specialists and collaboration with the AWMF (Association of the Scientific Medical Societies) in Germany and other European medical associations	n/a
Belgium	Decentralized with various organizations involved	EBMPracticeNet CEBAM (Belgian Center for Evidence-Based Medicine) KCE (Belgian Health Care Knowledge Center) College of Physicians	http://www.cebam.be/nl/Paginas/Home.aspx
Croatia	Centralized with various organizations involved and adapted locally	Ministry of Health of the Republic of Croatia (Ministarstvo Zdravlja Hrvatske) in collaboration with Croatian medical associations. Also, various hospitals and clinical centers adapting different guidelines to their standards	http://www.kbsd.hr/Klinicke-smjernice
Cyprus	Centralized but adapted locally	Health Insurance Organization (HIO) adopting clinical practice guidelines from the UK NICE and European medical associations	http://www.hio.org.cy/en/kko_intro.html
Czech Republic	Decentralized with various organizations involved	Czech Medical Association (CzMA) and other professional medical societies	http://www.cls.cz/english-info
Denmark	Centralized with various organizations involved	National Board of Health (Sundhedsstyrelsen) responsible for the development of national clinical guidelines (Nationale Kliniske Retningslinjer (NKR))	https://www.sst.dk/da/nkr
Estonia	Centralized with various organizations involved	Coordinated under the Estonian Health Insurance Fund (EHIF) in collaboration with the World Health Organization (WHO), the Medical Faculty at the University of Tartu (Ravijuhend)	http://www.ravijuhend.ee/juhendid/kasitlusjuhendid/
Finland	Centralized	National Institute for Health and Welfare (THL) Käypä hoito (Current Care) Unit Duodecim	http://www.kaypahoito.fi/web/english/guidelines/guideline?id=cCPG00004

Table 22.17 (continued)

Country	Coordination	Organization(s)	Website
France	Centralized with various organizations involved	Haute Autorité de Santé, HAS (French National Authority for Health)	http://www.has-sante.fr/portail/jcms/c_2036961/en/best-practice-guidelines
Germany	Centralized with various organizations involved	Working Group of the scientific medical societies (Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF)) and the Agency for Quality in Medicine (ÄZQ) National Disease Management Guideline Programme (Nationale VersorgungsLeitlinien (NVL))	http://www.awmf.org/en/clinical-practice-guidelines.html http://www.leitlinien.de/nvl/
Greece	Decentralized with various organizations involved	Various organizations involved. For example the University of Crete developed general practice and nursing guidelines.	http://www.greekphCPGuidelines.gr/
Hungary	Centralized with various organizations involved	The National Institute for Quality and Organizational Development in Healthcare and Medicines (GYEMSZI)	n/a
Iceland	Centralized	Directorate of Health (Velferdarraduneyti)	https://www.velferdarraduneyti.is/
Ireland	Decentralized with various organizations involved	Appointed by the Health Information and Quality Authority (HIQA), the National Clinical Effectiveness Committee (NCEC) develops National Clinical Guidelines which are implemented by the Health Service Executive (HSE)	https://www.hiqa.ie/areas-we-work/clinical-guideline-support http://health.gov.ie/national-patient-safety-office/ncec/national-clinical-guidelines-2-2/
Italy	Centralized with various organizations involved	National Guidelines System (Sistema Nazionale per le Linee Guida—SNLG) through the Health Ministry's General Directorate of Health Programming and the National Institute of Health (Istituto Superiore di Sanità—(ISS)	http://www.snlg-iss.it/home_en
Latvia	Centralized with various organizations involved	National Health Service (NHS) (Nacionālais veselības dienests (NVD))	http://www.vmnvd.gov.lv/lv/420-kliniskas-vadlinijas/klinisko-vadliniju-datu-baze
Lithuania	Centralized with various organizations involved	Ministry of Health in collaboration with medical faculties, the National Health Insurance Fund, the State Pharmaceutical Control Service, and the Mandatory Health Insurance Service	https://sam.lrv.lt/
Luxembourg	Centralized	Conseil Scientifique du Domaine de la Santé (Scientific Council)	http://www.conseil-scientifique.public.lu/fr/publications.html
Malta	Centralized with various organizations involved	Ministry for Health with input from the Mater Dei Hospital	https://health.gov.mt/en/Pages/health.aspx

Table 22.17 (continued)

Country	Coordination	Organization(s)	Website
Norway	Centralized with various organizations involved	Norwegian Directorate of Health responsible for the development, in consultation with Regional Health Authorities and clinical guidelines available at the Helsebiblioteket. No platform	http://www.helsebiblioteket.no/retningslinjer
Poland	Decentralized with various organizations involved	Clinical guidelines are formed by various medical societies and bodies and these are published online	n/a
Portugal	Decentralized with various organizations involved	Directorate-General for Health (Direcção-Geral da Saúde, DGS)	http://www.dgs.pt/directrizes-da-dgs/orientacoes-e-circulares-informativas.aspx
Romania	Centralized with various organizations involved	Ministry of Health of Romania (Ministerul Sanatatii)	http://www.ms.ro/ghiduri-clinice/
Slovakia	Decentralized with various organizations involved	National Institute of Quality and Innovations such as the General Practitioner Association (všeobecný praktický lekár)	http://www.vpl.sk/sk/guidelines_1/
Slovenia	Decentralized with various organizations involved	Slovenian Ministry of Health developed a handbook for clinical guidelines development but this is undertaken by various organizations such as the Slovene Medical Association	http://www.szd.si/
Spain	Centralized with various organizations involved	Guías de Práctica Clínica en el Sistema Nacional de Salud (Guides Clinical Practice in the National Health System)	http://portal.guiasalud.es/web/guest/guias-practica-clinica
Sweden	Centralized but adapted locally	National Board of Health and Welfare (NBHW) under the Ministry of Health and Social Affairs (Socialstyrelsen)	http://www.socialstyrelsen.se/nationalguidelines
Switzerland	Decentralized with various organizations involved	Various associations and professional groups involved	n/a
The Netherlands	Decentralized with various organizations involved	National Institute for Public Health and the Environment (RIVM), the Dutch Institute for Healthcare Improvement (CBO), the Dutch Council for Quality of Care, and the Dutch College of General Practitioners (NHG) involved with all guidelines available through one electronic platform	https://richtlijnendatabase.nl/
United Kingdom	Centralized	National Institute for Health and Care Excellence (NICE) for England and Wales with National Collaborating Centers (NCCs) and the Royal Colleges providing recommendations; Scottish Intercollegiate Guidelines Network (SIGN) for Scotland	https://www.nice.org.uk/ http://www.sign.ac.uk/

References

22.1 S.R. Hanney, M.A. González-Block: Health research improves healthcare: Now we have the evidence and the chance to help the WHO spread such benefits globally, Health Res. Policy Syst. **13**(1), 12 (2015)

22.2 D.M. Cutler, S. Kadiyala: The return to biomedical research: Treatment and behavioral effects. In: *Measuring the Gains from Medical Research: An Economic Approach*, ed. by K.M. Murphy, R.H. Topel (Univ. of Chicago Press, Chicago 2003) pp. 110–119

22.3 M.R. Macleod, S. Michie, I. Roberts, U. Dirnagl, I. Chalmers, J.P. Ioannidis, R.A.S. Salman, A.W. Chan, P. Glasziou: Biomedical research: Increasing value, reducing waste, Lancet **383**(9912), 101–104 (2014)

22.4 R. Grol, J. Grimshaw: From best evidence to best practice: Effective implementation of change in patients' care, Lancet **362**(9391), 1225–1230 (2003)

22.5 D.L. Sackett: Evidence-based medicine, Semin. Perinatol. **21**(1), 3–5 (1997)

22.6 O.M. Al-Quteimat, A.M. Amer: Evidence-based pharmaceutical care: The next chapter in pharmacy practice, Saudi Pharm. J. **24**(4), 447–451 (2016)

22.7 J.R. Hampton: Evidence-based medicine, opinion-based medicine, and real-world medicine, Perspect. Biol. Med. **45**(4), 549–568 (2002)

22.8 I. Skoglund, K. Segesten, C. Björkelund: GPs' thoughts on prescribing medication and evidence-based knowledge: The benefit aspect is a strong motivator: A descriptive focus group study, Scand. J. Prim. Health Care **25**(2), 98–104 (2007)

22.9 S. Every-Palmer, J. Howick: How evidence-based medicine is failing due to biased trials and selective publication, J. Eval. Clin. Pract. **20**(6), 908–914 (2014)

22.10 D. Blumenthal: Doctors and drug companies, N. Engl. J. Med. **351**(18), 1885–1890 (2004)

22.11 J. Abraham: *Science, Politics and the Pharmaceutical Industry: Controversy and Bias in Drug Regulation* (UCL, London 1995)

22.12 J. Hewitt-Taylor: Clinical practice guidelines and care protocols, Intensive Crit. Care Nurs. **20**(1), 45–52 (2004)

22.13 R. Grilli, C. Ramsay, S. Minozzi: Mass media interventions: Effects on health services utilisation, Cochrane Libr. (2002), https://doi.org/10.1002/14651858.CD000389

22.14 J.B. McKinlay: The promotion of health through planned sociopolitical change: Challenges for research and policy, Soc. Sci. Med. **36**(2), 109–117 (1993)

22.15 O.A. Arah: On the relationship between individual and population health, Med. Health Care Philos. **12**(3), 235–244 (2009)

22.16 A. Antonovsky: The salutogenic model as a theory to guide health promotion, Health Promot. Int. **11**(1), 11–18 (1996)

22.17 R.C. Brownson, J.F. Chriqui, K.A. Stamatakis: Understanding evidence-based public health policy, Am. J. Public Health **99**(9), 1576–1583 (2009)

22.18 T.B. Dinh-Zarr, D.A. Sleet, R.A. Shults, S. Zaza, R.W. Elder, J.L. Nichols, R.S. Thompson, D.M. Sosin: The task force on community preventive services: Reviews of evidence regarding interventions to increase the use of safety belts, Am. J. Prev. Med. **21**(4), 48–65 (2001)

22.19 B. Freeman, S. Chapman, P. Storey: Banning smoking in cars carrying children: An analytical history of a public health advocacy campaign, Aust. N. Z. J. Public Health **32**(1), 60–65 (2008)

22.20 L. Cornelsen, A. Carreido: *Health-Related Taxes on Foods and Beverages* (Food Research Collaboration, London 2015)

22.21 G. Weisz, A. Cambrosio, P. Keating, L. Knaapen, T. Schlich, V.J. Tournay: The emergence of clinical practice guidelines, Milbank Q. **85**(4), 691–727 (2007)

22.22 J. Miller: The scottish intercollegiate guidelines network (SIGN), Br. J. Diabetes Vasc. Dis. **2**, 47–49 (2002)

22.23 National Institute for Health and Care Excellence (NICE): Who we are, https://www.nice.org.uk/about/who-we-are

22.24 Haute Autorité de Santé: Contribute to the regulation by quality and efficiency: About HAS, http://www.has-sante.fr/portail/jcms/r_1455134/en/about-has (2013)

22.25 Agency for Quality in Medicine (ÄZQ): Nationale VersorgungsLeitlinien (NVL) (National Disease Management Guideline Programme), http://www.leitlinien.de/nvl/

22.26 Sistema Nazionale per le Linee Guida (SNLG) (National Guidelines System) and the Istituto Superiore di Sanità (ISS) (National Institute of Health): http://www.snlg-iss.it/home_en

22.27 Guías de Práctica Clínica en el Sistema Nacional de Salud (Guides Clinical Practice in the National Health System): http://portal.guiasalud.es/web/guest/guias-practica-clinica

22.28 J.M. Grimshaw, I.T. Russell: Effect of clinical guidelines on medical practice: A systematic review of rigorous evaluations, Lancet **342**(8883), 1317–1322 (1993)

22.29 J. Grant: Evaluating the outcomes of biomedical research on healthcare, Res. Eval. **8**(1), 33–38 (1999)

22.30 J. Grant, R. Cottrell, F. Cluzeau, G. Fawcett: Evaluating "payback" on biomedical research from papers cited in clinical guidelines: Applied bibliometric study, BMJ **320**(7242), 1107–1111 (2000)

22.31 G. Lewison, K. Wilcox-Jay: Getting biomedical research into practice: The citations from UK clinical guidelines. In: *Proc. 9th ISSI Conf., Beijing* (2003) pp. 152–160

22.32 G. Lewison, R. Sullivan: The impact of cancer research: How publications influence UK cancer clinical guidelines, Br. J. Cancer **98**(12), 1944–1950 (2008)

22.33 E. Pallari, A.W. Fox, G. Lewison: Differential research impact in cancer practice guidelines' ev-

idence base: Lessons from ESMO, NICE and SIGN, ESMO Open **3**(1), e000258 (2018)
22.34 D. Kryl, L. Allen, K. Dolby, B. Sherbon, I. Viney: Tracking the impact of research on policy and practice: Investigating the feasibility of using citations in clinical guidelines for research evaluation, BMJ Open **2**(2), e000897 (2012)
22.35 S.M. Noar: A 10-year retrospective of research in health mass media campaigns: Where do we go from here?, J. Health Commun. **11**(1), 21–42 (2006)
22.36 D.P. Phillips, E.J. Kanter, B. Bednarczyk, P.L. Tastad: Importance of the lay press in the transmission of medical knowledge to the scientific community, N. Engl. J. Med. **325**, 1180–1183 (1991)
22.37 G. Lewison, S. Tootell, P. Roe, R. Sullivan: How do the media report cancer research? A study of the UK's BBC website, Br. J. Cancer **99**(4), 569–576 (2008)
22.38 World Health Organization (WHO): Health statistics and information systems—Estimates for 2000–2012, http://www.who.int/healthinfo/global_burden_disease/ estimates/en/index2.html (2012)
22.39 G. Lewison, G. Paraje: The classification of biomedical journals by research level, Scientometrics **60**(2), 145–157 (2004)
22.40 M. Begum, G. Lewison: Web of science research funding information: Methodology for its use in analysis and evaluation, J. Scientometr. Res. **6**(2), 65–73 (2017)
22.41 V. Bakare, G. Lewison: Country over-citation ratios, Scientometrics **113**(2), 1199–1207 (2017)
22.42 E. Pallari, G. Lewison, R. Sullivan: How is chronic non-communicable respiratory conditions research reported in European newspapers? An impact assessment for policy, Clin. Respir. J. **11**(5), 657–665 (2017)
22.43 A.D. Purushotham, G. Lewison, R. Sullivan: The state of research and development in global cancer surgery, Ann. Surg. **255**(3), 427–432 (2012)
22.44 T.G. Weiser, S.E. Regenbogen, K.D. Thompson, A.B. Haynes, S.R. Lipsitz, W.R. Berry, A.A. Gawande: An estimation of the global volume of surgery: A modelling strategy based on available data, Lancet **372**(9633), 139–144 (2008)
22.45 M. Begum, G. Lewison, J.S.F. Wright, E. Pallari, R. Sullivan: European non-communicable respiratory disease research, bibliometric study outputs funding, PLoS ONE **11**(4), e0154197 (2016)
22.46 R. Rajendram, G. Lewison, V.R. Preedy: Worldwide alcohol-related research and the disease burden, Alcohol Alcohol. **41**(1), 99–106 (2006)
22.47 E. Pallari, G. Lewison, O. Ciani, R. Tarricone, S. Sommariva, M. Begum, R. Sullivan: The impacts of diabetes research from 31 European Countries in 2002 to 2013, Res. Eval. **27**(3), 270–282 (2018)

Elena Pallari
Centre for Implementation Science, Health Service & Population Research Department
King's College London
London, UK
elena.pallari@kcl.ac.uk

Elena Pallari is a Research Assistant at King's College London. She previously held positions on a wide range of multidisciplinary projects in the conduct, evaluation and evidence-synthesis of clinical studies at the University College London Medical School, the Brunel Institute of Bioengineering and Imperial College London Business School. Elena also works as a Life Sciences Consultant for various pharmaceutical clients, private firms and research groups.

Grant Lewison
Division of Cancer Studies
King's College London
London, UK
grantlewison@aol.co.uk

Grant Lewison is a Senior Research Fellow at King's College London. He has worked in bibliometrics for nearly 30 years and he is the Director of Evaluametrics Ltd., a bibliometric consultancy. His main interests are in alternative measures of biomedical research impact, such as citations on clinical guidelines and in the mass media. Previously, he worked at University College London, City University, and the Wellcome Trust.

23. Societal Impact Measurement of Research Papers

Lutz Bornmann, Robin Haunschild

What are the results of public investment in research from which society actually derives a benefit? The scope of research evaluations becomes broader when societal products (outputs), societal use (societal references), and societal benefits (changes in society) of research are considered. This chapter presents an overview of the literature in the area of societal impact measurement of scientific papers. It describes major research projects on societal impact measurements. Problems of societal impact assessments are discussed as well as proposals to measure societal impact. The chapter discusses the role of alternative metrics (altmetrics) in measuring societal impact. There is an ongoing debate in scientometrics as to whether altmetrics are able to measure this kind of impact.

In the second half of the twentieth century, it was initially assumed in the science policy of many countries that a society benefitted most from a science that was oriented to its own success criteria. *Godin* and *Dore* [23.1, p. 1] concluded:

> Although scientific policy has for a time been driven by the 'policy for science' philosophy or ideology, there has never been any doubt in the minds of policy-makers that the ultimate aim for funding science and technology was socio-economic goals such as national security, economic development, welfare and the environment.

Whereas scientific products and activities (such as manuscripts, research projects, or programs) were initially evaluated almost exclusively by means of the peer review procedure [23.2], later quantitative indicators were also used for research evaluation, mainly bibliometric indicators (based on publication output and impact) [23.3]. In this period, in which science was primarily judged by the benefit for itself, the societal and economic relevance of research was hardly doubted and taken as given [23.4]. An economically successful society needs research which—judged by the criteria inherent in science—operates at the highest level.

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_23

Since the 1980s, science policy has increasingly departed from this paradigm [23.5]. Just like other areas of society, science is increasingly influenced by the effects of an audit society [23.6] in which the use of public funds requires comprehensive accountability (keyword *new public management* [23.7]). However, it is no longer sufficient for accountability just to demonstrate the originality and excellence of research in self-regulated procedures [23.8], but the societal relevance of research, the practical real-world benefits from research, and the fulfilment of societal needs must also be shown [23.9–11]. According to *Godin* and *Dore* [23.1, p. 5]:

> What one expects today is measures of the impact of science on human lives and health, on organizational capacities of firms, institutional and group behavior, on the environment, etc.

There is a special focus on the production of direct societal and economic utility and impact [23.12].

This book chapter provides an overview of the literature on societal impact measurements of research. The chapter focuses on studies that analyze and discuss impact measurements on society—setting aside the literature dealing with forms of interaction between science and society other than impact. In the development of indicators for the measurement of societal impact, *Donovan* [23.13] distinguishes three phases, where it is particularly the last phase that is dealt with in this book chapter:

1. Technometrics (capture of data on investments from industry in science, commercialization of scientific products, and technology transfer)
2. Sociometrics (mapping of research outcomes onto existing government social statistics)
3. Case studies (demonstrating the societal impact of research).

However, the use of case studies for the measurement of societal impact forms only one section in this book chapter (Sect. 23.3.1). Prior to this, Sect. 23.1 outlines why it is increasingly regarded as necessary to measure the societal impact of research activities, how societal impact is defined, and which problems there are with societal impact measurement. Section 23.2 examines current practice in societal impact measurement in various areas (with funding bodies, in national evaluation exercises as well as with multidimensional frameworks and productive interactions). Starting from a critical discussion of the methods currently favored for societal impact measurement—the case studies—Sect. 23.3 discusses the use of quantitative indicators for societal impact measurement.

In classical scientometrics, scholarly communication predominantly formed the general framework of impact measurement. Impact mostly means perception by scientists as measured by citations, although some evaluation studies had already left this framework. For example, *Grant* [23.14] dealt with measures for the impact of medical research on clinical management and the general public. Leaving the scholarly communication framework along with a broadening of the notion of impact towards societal impact led to increased complexity in scientometrics. The field faced the more general questions of *impact on whom* [23.15] and *heterogeneity of users* regarding the so-called *alternative metrics* (altmetrics). Altmetrics count tweets, posts, and many other mentions of scholarly papers on the web. These questions not only affect the conceptual foundations of impact, but also the methodological implications, and the technical challenges of possible measurements of societal impact.

Thus, Sect. 23.4 in this book chapter covers the use of altmetrics for the measurement of societal impact. Altmetrics are regarded as a promising possibility for measuring the societal impact of research. The section outlines how societal impact measurement could be performed using altmetrics (chiefly with social media metrics, citations in patents, citations in clinical guidelines, and mentions in policy-related documents). Here, two key aspects are (1) the time- and field-normalization of altmetrics (with similar methods as for citations) and (2) the target-oriented measurement of impact. For example, the impact measurement of papers on policy-related documents may be a target-oriented impact measurement of research on the area of politics.

23.1 Definition of Societal Impact as Well as Reasons for and Problems with the Measurement

This section deals with three basic topics: reasons for societal impact measurements, definition of societal impact, and problems with societal impact measurement.

23.1.1 Reasons for Societal Impact Measurements

Until the 1980s, science could largely act autonomously in many countries, as long as it remained oriented towards its own excellence and originality criteria. Today, there "is a growing interest in methods of evaluation that focus on (or include) the societal relevance of research" [23.16, p. 20], which involves a weakening of the separation of fundamental categories, like private and public sectors, science and values, producers and users of knowledge [23.17]. In what follows, we would like to discuss the main reasons which have led to increased interest in the measurement of the societal impact of research in recent decades.

Accountability in an Audit Society

In an era in which governments are seeking to reduce spending by any means possible, even areas of society whose importance used to be (or is) unquestioned hung in the balance [23.3]. These areas—which include science—increasingly have to demonstrate that they provide a considerable contribution to society and the public investments are justified [23.18]. The audit society [23.19] would like to use societal impact measurement to obtain information on the way particular pieces of research (especially fundamental research) actually makes an important contribution [23.4, 8].

Societal Relevance of Research

The publication of research results (in high-impact journals) may be very important for scientists, but society does not benefit until the results are applied or used (e. g., via guidelines, medicaments, diagnostic tools, machines, and devices) [23.20–22]. *Ernø-Kjølhede* and *Hansson* [23.23, p. 132] concluded:

> Universities are to an increasing degree subject to a requirement to conduct research in close interaction with users and other stakeholders, implying that the research must cut across scientific disciplines and theories and focus on problem-solving and practical use-value.

In this connection, *Cohen* [23.24] describes an interesting conflict over more than $10 billion, which rich countries have poured into programs for malaria control. Researchers at the Center for Global Development (CGD) in Washington, D.C., have questioned the usefulness of this funding: did it really prevent malaria cases and save lives? These and similar questions illustrate a trend towards a search for reliable data about the benefits of research (e. g., in medicine the number of lives actually saved by the billions invested).

Strengthening the Economy of a Country

It is one of the most important objectives of the successful policy of a country to enhance international industrial and economic competitiveness [23.13]. According to *Toronto Region Research Alliance* [23.25, p. 3]:

> Modern economies increasingly view science and technology (ST) as a major competitive edge and are exploring strategies to use innovation and intellectual capital to drive economic development and growth.

Science policy therefore increasingly demands that contributions from science, particularly to strengthening the economy of a country, are demonstrated.

Proof of Various Kinds of Usefulness

In a perfect world of science, it should actually be the case that every research activity at a university is of high quality and at the same time produces societal benefits [23.26]. But in reality, universities have very diverse orientations and profiles which result in one case in a high quality of research for research (in particular subject categories) and in another case to great usefulness of the research for other areas of society. Thus, there is a broad spectrum of research and utility which "needs a holistic approach that examines the main channels that bind universities to the rest of society" [23.27, p. iv]. Whereas conventional impact measurement could (and can) chiefly demonstrate the usefulness of research for research, broad impact measurement should be in a position to demonstrate usefulness which was hardly possible with conventional impact measurement [23.18, 28]. An example of an obvious area for universities is the educational function provided for the society.

Focus on Relevant Research

For several years, a general trend has been evident in science policy to increase the application of research funds to tackle quite particular societal issues (such as climate change, social cohesion, and globalization) [23.12, 29] and a country's needs [23.30]. Societal impact measurement should on the one hand enable the universities to prove their engagement with the issues and needs. The use of particular indicators in the measurements should, on the other hand, create incentives for engagement with particular issues and needs [23.11].

23.1.2 Definition of Societal Impact

Before we present the definition of societal impact by *Wilsdon* et al. [23.31], we would first like to present some examples from the literature which show what can be understood concretely by the societal impact of research. We would like to begin with the paper by *Dance* [23.32], in which the following examples for broad impact of research are mentioned [23.32, p. 398]:

> For example, ... an engineer might chat to local industry figures about their environmental concerns, and work out how to use academic inventions to solve their problems. University knowledge—or technology—transfer offices may be able to help scientists to forge relationships with industry partners.

According to *Wilsdon* et al. [23.31] societal impact of research can look like this [23.31, pp. 44,45]:

> References to, citations of or discussion of an academic or their work; in a practitioner or commercial document; in media or specialist media outlets; in the records of meetings, conferences, seminars, working groups and other interchanges; in the speeches or statements of authoritative actors; or via inclusions or referencing or web links to research documents in an external organization's websites or intranets; in the funding, commissioning or contracting of research or research-based consultancy from university teams or academics; and in the direct involvement of academics in decision-making in government agencies, government or professional advisory committees, business corporations or interest groups, and trade unions, charities or other civil society organizations.

As the examples of *Dance* [23.32] and *Wilsdon* et al. [23.31] show, there is a broad spectrum of possible societal impact of research. An important reason for the variety could be that a range of social factors play an important role in the attribution of societal impact. According to the results of *Samuel* and *Derrick* [23.33], who conducted interviews with evaluators from the research excellence framework (REF) [23.33, p. 237]:

> The possibility of impact being realized [is]... more related to a range of social factors, than adequately reflecting the nature of the research, or the efforts of the researchers themselves. This concept is built from the observation that the societal impact of science is not value-free and neutral.

Correspondingly, there is also no uniform standard definition of societal impact. Whereas science generally uses impact to mean citation impact [23.34], there is no clear understanding for impact in society. "Impact is still a contested term, with a variety of definitions and understandings of its implications" [23.31, p. 44].

A very generally framed definition of societal impact is that of *Wilsdon* et al. [23.31, p. 6]:

> Research has a societal impact when auditable or recorded influence is achieved upon non-academic organization(s) or actor(s) in a sector outside the university sector itself—for instance, by being used by one or more business corporations, government bodies, civil society organizations, media or specialist/professional media organizations or in public debate. As is the case with academic impacts, societal impacts need to be demonstrated rather than assumed. Evidence of external impacts can take the form of references to, citations of or discussion of a person, their work or research results.

Societal impact measurement is therefore a matter of every measurable influence (effect) of research which can be demonstrated outside of research in a particular sector of society [23.35]. The *RQF Development Advisory Group* [23.36] (RQF is an abbreviation for research quality framework) and *Samuel* and *Derrick* [23.33] cite the following areas in which one could expect impact from research: informed public debate and improved policy-making (social benefit), adding to economic growth and wealth creation (economic benefit), improved management of natural resources (environmental benefit), supporting greater understanding of where we have come from, and who and what we are as a nation and society (cultural benefit).

Against the backdrop of the definition of *Wilsdon* et al. [23.31] it is astonishing to see that the majority of interviewees in the study by *Samuel* and *Derrick* [23.33] (see above) defined societal impact as an *outcome*, where this then led to a *change* or a *difference* in society. The results of the study could indicate that the interviewees often do not clearly distinguish output, outcome, impact, and effects. We would define these terms as follows: Whereas the outcome is a matter of a concrete research result (e. g., unemployment leads to psychosomatic ailments), output is a concrete product resulting from the research result (e. g., a publication or a lecture). Impact then describes those events, products or changes (e. g., in clinical guidelines), which are based on the outcome (in a sector of society). If this impact leads to broad, general societal changes (e. g., important political decisions), we can speak of effects [23.1, 18].

The measurement of outcome, output, impact, and effect can be further refined with respect to various parameters, e. g., temporal aspects. Thus, impact can be measured broadly, or it can be investigated whether particular research activities meet a particular social need [23.37]. According to *Lähteenmäki-Smith* et al. [23.18, p. 35]:

> It is useful to divide impacts as follows: anticipated and unanticipated; inside and outside the target area (or relevant or irrelevant); productive and detrimental (or neutral in impact).

Outcome, output, impact, and effect can additionally be differentiated on a temporal scale, as demonstrated in what follows taking the example of the outcome [23.9]:

> We can thus differentiate between immediate (e. g., publications, prototypes), intermediate (e. g., partnership-based cooperation, new/improved products) and ultimate (e. g., improved industry competitiveness) outcomes [23.18, p. 34]

According to *Ruegg* and *Feller* [23.38] it is the longer term outcomes which are of interest socially (e. g., improvements in the quality of life).

23.1.3 Problems with Societal Impact Measurement

As we can see from the difficulties with the definition of societal impact, the phenomenon of *societal impact* has a much more complex structure than the phenomenon of *academic impact*. For example, whereas the target group for academic impact is relatively clearly defined (the active scientists), with societal impact it is diffuse: it is a matter of all groups of people and areas which do not belong to science. The breadth alone that has to be taken into account in the measurement of societal impact raises many questions about the feasibility of the measurement and the interpretation of the results. The studies concerning the measurement of societal impact in recent years have mentioned numerous problems associated with this measurement. In what follows, we would like to outline a selection of the most important problems.

Various Exploitation Channels

We can generally assume that benefits of research for society arise through various "exploitation channels" [23.39]. There is the danger of concentrating on too few channels in the measurement of impact (in the worst case, on only one) and selecting indicators which use the channels that are the simplest to measure. This introduces a bias in the measurement of impact which raises questions about a reliable and valid impact measurement. We can additionally assume that the value of scientific research is assessed differently by stakeholders in society [23.40]:

> Scientific discoveries and inventions affect many different aspects of society. Different decision-makers will make different evaluations of the same outcomes. It is thus methodologically impossible to summarise the impact of scientific knowledge and inventions in any single indicator.

Attribution of Societal Impact

If research has effects in a societal sector, it is generally difficult to determine what portion of benefits should be attributed to a certain piece of research. This means that societal impact measurement has a causality or attribution problem [23.39]: which research caused a particular effect in society? For example, with societal impact like the improved state of public health, it is almost impossible to attribute this impact to particular pieces of research [23.5]. According to *Milat* et al. [23.11] "research impacts are complex, non-linear, and unpredictable in nature". The benefits of research are often "subtle, heterogeneous, difficult to track or measure, and mostly indirect" [23.41, p. 528]. They are derived "from serendipitous findings, good fortune, and complex networks interacting and translating knowledge and research" [23.35]. Thus, the paths from research to impact generally remain a black-box [23.1, 42].

There are often several scientists from various projects, research groups, funding schemes or funding sources or countries working on a research topic which makes it difficult to attribute the impact of the research to particular units [23.43, 44]. *Nightingale* and *Scott* [23.21, p. 543] concluded:

> The benefits of research are complex, contingent, often indirect, and typically involve the generation of intellectual infrastructure and the transfer of trained people, new instrumentation and methods, tacit skills, and networks.

Khazragui and *Hudson* [23.3] regard it as a fundamental problem with the impact measurement of research in the area of policy that the policy tends to be based upon a large body of research efforts which have constituted *the commons* [23.45].

Time Lag

The time lag between the outcome of the research and the later impact varies enormously: in one case the impact is quickly visible, in another case very late [23.43, 44]. This also involves the problem that the time of the impact measurement is hard to determine [23.27], and the result of the impact measurement is highly dependent on the selected time lag [23.46]. The length of the time lag should ultimately be oriented to the complexity of the object being researched. It should generally be borne in mind with societal impact measurement that one can only make positive and not negative statements about impact.

For *Milat* et al. [23.11] it is the long-term societal impact measurement of research which is most problematic, since this can involve periods of 15 years and more:

> It has been observed that, on average, it takes over 6 years for research evidence to reach reviews, papers, and textbooks, and a further 9 years for this evidence to be implemented into practice.

However, the literature also mentions shorter time frames for the measurement of the longer term impact: thus *Khazragui* and *Hudson* [23.3] for example mention a time frame of 6 years, related to the commercialization of research findings.

But research evaluations generally aim to measure (citation) impact over a shorter time window. Only up-to-date data forms a relevant basis for decision-making [23.43]. However, there is a risk that research bringing short-term benefits is over-emphasized [23.39].

Social Factors

University and nonuniversity research institutions have various profiles and missions which depend on internal and external institutional contexts and determine the particular research guidelines [23.44, 47]. According to *Molas-Gallart* et al. [23.27, p. 8] "there is no one model of the successful university." Therefore, only university and nonuniversity research institutions from the same context should be compared with one another. For this it is important that evaluations are performed as often as possible to provide space for various possibilities of impact measurement [23.48, p. 7]:

> The impact to be 'demonstrated' could be that of a project or research unit, of a program, of a funding body/strategy, of an area of research, or of the research system as a whole—each captured at different points in time, and relative to varying time horizons and to different types and methodologies of research.

These possibilities of impact measurement should also be utilized in practice to avoid leaving the impact of some research "under the radar" [23.49].

The results of the study of *Samuel* and *Derrick* [23.33] show that it is not only the social factors surrounding the production of research results which should be considered with impact measurement, but also social factors surrounding impact measurement [23.33, p. 237]:

> A number of interviewees ($n = 18$) recognized that impact was contingent on social processes. They perceived that the possibility of impact being realized was more related to a range of social factors, than adequately reflecting the nature of the research, or the efforts of the researchers themselves. This concept is built from the observation that the societal impact of science is not value-free and neutral and that science does not have an impact based solely on its particular capabilities. Rather, scientific research, its shaping and development, and its application to society, is related to multilayered social factors.

Serendipity

The results of research are generally hard to predict and serendipity plays a role in the research process which should not be underestimated [23.11, 50, 51]. This also makes the impact measurement difficult [23.27, 52]. Particularly in relation to the long-term impact of research the possibility of serendipity and unpredictable effects should always be considered [23.44, p. 69]:

> Is it possible that scientists who laid the groundwork for Google or wireless communication or their peers, or any metrics available today for that matter, could have predicted the multi-million dollar value of their original work? Is it possible to predict which projects undertaken today will lead to unfathomable transformations in the lives of future generations? Will metrics help protect seemingly obscure projects that could one day hold the key to these transformations, or will they encourage their dismissal?

The latter should be prevented if possible.

Societal Quality

Citations are often used as a proxy for quality in bibliometrics. This means that one actually wants to measure the quality of research and uses data for this which can measure the construct *quality*. According to *Martin* and *Irvine* [23.53] citations measure a specific part of quality, that is impact. Two other parts of quality can scarcely be measured by citations: importance and correctness of research. In discussions of the measurement of societal impact it is noticeable that the important connection to the quality of the research is lacking. Only occasionally do references to this occur in the literature: For *van der Meulen* and *Rip* [23.45] "it is not clear how to evaluate societal quality, especially for basic and strategic research" [23.25, p. 11]. According to *de Jong* et al. [23.4, p. 62]:

> Evaluating societal quality suffers from methodological problems, as it is difficult to attribute impact to specific inputs: The relation between knowledge and impact is complex and innovations are based on a variety of (knowledge) sources.

An impact measurement which has no relation to the quality of research should be avoided as far as possible in research evaluation. For example, we can assume an immense impact outside of science (such as in the media) especially for scientific activities which are ethically questionable (such as misconduct or fraud). In order to avoid giving scientists false incentives by the use of metrics (e. g., number of mentions on the news), the quality (correctness) of the research connected with impact measurement should always stand in the foreground. An impact measurement should therefore only be undertaken for publications which fulfil the scientific standards in their specialist area (for example, by including in the analysis only publications accepted for publication by a peer review procedure).

Research with Negative and Positive Effects

Research can have both positive as well as negative impact on society [23.44], as the following example makes clear [23.54, p. 6] clarifies:

> Environmental research that leads to the closure of a fishery might have an immediate negative economic impact, even though in the much longer term it will preserve a resource that might again become available for use. The fishing industry and conservationists might have very different views as to the nature of the initial impact—some of which may depend on their view about the excellence of the research and its disinterested nature.

In such cases, it is difficult to undertake a societal impact measurement which is oriented to the benefit for the society.

Subject-Specific Differences

There are clear subject-specific differences as to which societal impact one can expect and the possible extent of this societal impact [23.32, p. 398]:

> The type of broader-impact project can differ between disciplines. A mathematician could explain his or her research to scientists in other fields, who might find it useful for modelling their own systems, suggests John Hand, head of impact at the UK Engineering and Physical Sciences Research Council. Engineers, by contrast, might offer applied projects with more direct practical impact, such as ways to scale up production processes.

The specifics of the specialist areas lead—as with citation impact—to the societal impact from different specialist areas hardly being comparable with one another.

23.2 Societal Impact Considerations in Evaluative Practice

In this section, the following questions are addressed:

- How is societal impact assessed at funding bodies?
- How is societal impact measured in national evaluation systems?
- Which frameworks have been proposed for societal impact measurements?
- What is understood by *productive interactions* as an alternative to impact measurements?

23.2.1 Societal Impact Assessments at Funding Bodies

A number of funding bodies have in recent years begun to consider not only the expected scientific outcome of research in the allocation of research funding, but also the possible impact of the planned research beyond the area of science. There are numerous references in the literature to funding bodies which consider the expected broader impact in the selection process. Examples of these are the US National Science Foundation (NSF), the Research Councils of the UK (RCUK), the Research Council of Norway (RCN), and the EU Framework Program/Horizon 2020 (EU FP7/H2020) [23.43, 55]. According to *Dance* [23.32, p. 397]:

> The interest in broader impact is rising. In 2009, the seven government-funded granting agencies that make up Research Councils UK (RCUK) began requiring applicants to delineate their impact plans. The Swiss National Science Foundation (SNSF) added a section on broad impact to its application forms in 2011. The US National Science Foundation (NSF) has long required applicants to combine scientific value with impact outside the lab, and in 1997 made broader impact an explicit part of the grant review. The foundation started requiring a separate section on impacts in applications this year.

In the UK, measures to assess ex ante societal impact are commonly recognized as the "pathways to impact" statement [23.33].

The consideration of societal impact in the evaluation of research applications has a number of consequences for the peer review process:

1. In the peer review process, research is assessed by specialist colleagues [23.9]. Since the specialist colleagues generally cannot or do not wish to estimate the societal impact of a project [23.56], the question arises of involving other actors than scientists working in the same area [23.45]: Which other actors could be included in the assessment process in order to assess the possible societal impact of the applications? Should their opinion—if they are included—be considered equal to that of the academic actors [23.43]?
2. It is not only the effort for the assessment of the proposals that is increasing, but also the effort for the applicants. The applicants not only have to consider the possible scientific outcome of a project, but also the possible outcome beyond science [23.9, 22].
3. Since the assessment of the scientific quality of applications involves a number of criteria (such as the importance or methodological rigor of an application), attention must be paid to the relevant criteria for the assessment of societal impact [23.21].

23.2.2 National Evaluation Systems and the Measurement of Societal Impact

In recent years, several countries have established national research assessment exercises, with which a comparative evaluation of research institutions within a country is performed. In the framework of these exercises, the institutions are generally evaluated not only by criteria immanent in science, but also by the usefulness of their research for society, where the economic value of publicly funded research stands in the foreground [23.13]. According to *Ovseiko* et al. [23.55]:

> In 1989, the UK was the first country in the world to implement a performance-based research funding system, the research assessment exercise (RAE), now the research excellence framework (REF), and since then at least thirteen more countries, including Australia, New Zealand, Hong Kong (China), and several EU countries, have introduced such systems.

In the following, we would like to present three examples of these national research assessment exercises.

The best-known example of a national evaluation system, and which was the first to look at the impact beyond science, is the UK REF [23.23]. A detailed description of societal impact measurement in the REF can be found in *Samuel* and *Derrick* [23.33]. The aims of the REF are [23.55]:

> Primarily to provide a basis for resource allocation, accountability for public investment in research, and benchmarking information and reputational yardsticks for the higher education sector.

The term societal impact is very broadly defined in the framework of the REF: it is a matter of the effects/benefits of the research for the economy, society, culture, public policy or services, health, the environment, or quality of life [23.57]. The evaluation of the societal impact of an institution is directly linked to its funding allocation in the REF. The criterion makes up 20% of the overall assessment of an institution, on which basis the funding allocation is decided.

Since the mid-1990s, the Standard Evaluation Protocol has been published in the Netherlands, with a description of the methods that should be used for the six-yearly evaluation of the Dutch research institutions. The protocol currently used refers to the period between 2015 and 2021 (https://www.knaw.nl/nl/actueel/publicaties/standard-evaluation-protocol-2015-2021). In evaluations in the Netherlands, not only the quality of the research is assessed, but also its social, economic, and cultural relevance for areas of society outside science. Here institutions should show how far their research affects specific stakeholders or specific procedures in society (such as laws or regulations). The methods that are used for the evaluation of the societal relevance of research were developed in the project Evaluating Research in Context (ERiC) [23.29].

In Australia, the Australian Research Council organizes the national research evaluation framework, called Excellence in Research for Australia (ERA). In this framework, research is evaluated with the help of the peer review procedure and selected indicators. The indicators used for the evaluation (for example bibliometric indicators), have been identified or developed in collaboration with the relevant specialist bodies. According to the information of *Wilsdon* et al. [23.31, p. 128]:

> In the Australian ERA, some quantitative measures of broader impact are assessed, namely patents, plant breeders' rights, registered designs and research commercialisation income. We heard evidence that there are concerns with this approach, related to the implied narrow definition of societal impact and the potential that focusing on a small number of metrics might significantly skew behaviour.

Particular attention is applied in the evaluation to the commercialization of research discoveries [23.58]. It is generally simpler to measure than other effects.

23.2.3 Frameworks for the Measurement of Societal Impact

In recent years, a number of more complex frameworks have been developed for measuring the quality of research and its societal impact. An overview of these frameworks can be found, for example, in *Bornmann* [23.59, 60] and *Milat* et al. [23.11]. Even if the frameworks differ in their terminology and approaches, the shared common features include [23.11]:

> Assessment of traditional research outputs, such as publication and research funding, but also a broader range of potential benefits, including capacity, building, policy and product development, and service development, as well as broader societal and economic impacts.

We would like to present two frameworks as examples in what follows:

1. In the framework of the Dutch evaluation of research with the standard evaluation protocol, for example, the research embedment and performance profile (REPP) approach was developed [23.16]. For this a number of social domains were defined, in which scientists operate:

   a) Science and certified knowledge
   b) Education and training
   c) Innovation and professionals
   d) Public policy and societal issues
   e) Collaboration and visibility.

   For each domain, a variety of criteria and indicators were developed (such as production of qualified researchers or migration of researchers to positions in business organizations), with which research can be evaluated. For the presentation of the results, a special graphical layout was developed, where the values for the individual indicators are represented in a radar graph. These radar graphs provide a comprehensive visual depiction of the research of a unit (e. g., of an institute).
2. One of the best-known frameworks with which the societal impact of research can also be measured, is the payback framework [23.61]. This framework was originally developed for the investigation of *impact* or *payback* of health services research [23.62]. The framework is a research tool, with which data on research impact of a research unit can be compiled, in order then to make a comparative cross-case analysis. For this purpose, the framework provides a multidimensional categorization of benefits from research which refers to the following five areas:

   a) Knowledge
   b) Benefits to future research and research use
   c) Political and administrative benefits
   d) Health sector benefits
   e) Broader economic benefits.

   The data for the five categories can be collected from surveys of decision-makers and analysis of documents [23.9]. In recent years, the framework has been used in a number of different contexts, which extend beyond the borders of the health services [23.63].

Not only in the frameworks presented here, but also in others, *Milat* et al. [23.11] see a significant problem in maintaining the right balance between comprehensiveness and feasibility. In order to represent quality of research multidimensionally, a manifold of indicators is necessary. But the more indicators a framework uses, the more difficult it is to install it in an institution and the more awkward it is for concrete application in research evaluation.

23.2.4 Productive Interactions

In Sect. 23.1.3 we described a number of problems that are connected to societal impact measurement. One example of these problems consists in the often late appearance of the societal impact of research. However, in the evaluation of research a prompt societal impact measurement is desirable. Against the backdrop of this and other problems with societal impact measurement, some authors have suggested dropping impact measurement and focusing instead on productive interactions of researchers with societal stakeholders [23.64]. The background for this change in perspective consists in the reasonable assumption that the productive interactions will later also lead to societal impact [23.65]. As has been shown by the results of two case studies in the area of law and architecture, productive interactions are an [23.4, p. 70]:

> Important way of circulating knowledge between science and society. The intensity of the collaboration informs us about the type and amount of knowledge that is circulated. On top of that, collaborations are an indication of societal quality.

With the change of societal impact measurement to the assessment of productive interactions, one takes a glance at the efforts undertaken by an institution to achieve societal impact [23.27]. According to *SIAMPI* (Social Impact Assessment Methods for research and funding instruments, see http://www.siampi.eu) [23.66, p. 2]:

> Interactions are determined to be 'productive' when they lead towards changes in behaviour on either side—that is with researchers (changes in research agenda) or with stakeholders (changes in behaviour).

The productive interactions can consist of personal interactions (e. g., through video-conferencing), through some kind of material *carrier* of the interaction (e. g., written means of communication, exhibitions, models, or films), or through economic exchange between researchers and potential stakeholders (financial interactions). The interactions can then take place when the research agenda is specified, during the research itself or after the end of the research project [23.29]. The intensity of the interaction can vary depending on the context between "very incidental and informal relations to highly organized and professionalized networks" [23.66, p. 5]. In general, an increased intensity of interactions allows the assumption of a higher probability of societal impact [23.67]. A good example of a productive network which is intended to accelerate the propagation of knowledge and technology from academia down to industry is the knowledge transfer networks (KTN, https://connect.innovateuk.org/knowledge-transfer-networks).

For the intensification of productive interactions, *Bornmann* and *Marx* [23.68] have suggested that scientists write what are known as assessment reports to summarize the research status in a particular research area or on a particular research topic (e. g., in climate change research). These reports should be written so as to be comprehensible for people outside of the specialist area or for people outside of science.

23.3 Case Studies and Quantitative Indicators

As shown in Sect. 23.2.3, some frameworks for societal impact measurement have already been developed. However, the design of the frameworks is generally very complex, which complicates their use. The measurement of productive interactions is a promising method for the estimation of societal impact; but ultimately the impact is not directly measured. Since for societal impact measurement no generally satisfactory method has yet been found, scientometrics is in search of indicators which are more or less simple to use. Before we present some promising possibilities in Sect. 23.4 for the reliable and valid measurement of societal impact, we would first like to discuss the currently preferred method for societal impact measurement in Sect. 23.3.1: case studies. Criticism of the use of case studies (Sect. 23.3.1) led to awareness of the necessity of the development (use) of quantitative indicators for societal impact measurement (Sect. 23.3.2) [23.1].

Currently, it is chiefly alternative metrics (altmetrics) which are seen as a possibility for measuring the societal impact of research quantitatively [23.69]. Since this area has been accorded ever greater importance in recent times and has developed into a research area of its own in scientometrics, this chapter deals with altmetrics in Sect. 23.4. There we will—besides the social media metrics (such as tweets or readers) in Sect. 23.4.1 and references to publications in policy-related documents in Sect. 23.4.4—chiefly discuss two metrics which according to *Wilsdon* et al. [23.31] appear particularly promising for societal impact measurement [23.70, p. 49]:

> In sum, while some alternative metrics seem to reflect types of impact that are different from that of traditional citations, only Google patent citations and clinical guideline citations can yet be shown to reflect wider societal impact.

23.3.1 Case Studies—Advantages and Disadvantages

Against the backdrop of the difficulty of measuring societal impact quantitatively or with indicators, the method of case studies is currently preferred for the demonstration of societal impact (with institutions). Especially in connection with the UK REF, case studies are repeatedly discussed in the scientometric literature; but they also appear in other evaluation contexts [23.71]. In research evaluations, case studies are used to demonstrate success stories or best practices in the achievement of societal impact through research which was performed in a particular institution [23.49]. These case studies are generally produced in the course of self-evaluation and examined by review panels in the external evaluation [23.43].

The UK REF stipulates to the institutions an impact template for the recording of the case studies, which should cover about four sides. According to *Salter* and *Martin* [23.41, p. 232]:

> Within these templates, institutions had to nominate pieces of underpinning research conducted at their institutions—for example, reports in the grey literature, or academic journal articles—and explain how this research had had an 'impact' on society.

The research selected for the case studies should have a quality level as high as possible—judged by scientific standards. A case study in the UK REF should have five sections [23.3, p. 56]:

1. A summary
2. A description of the underpinning research
3. The references
4. The impact
5. Corroborating evidence for this impact.

Some examples of case studies from the UK REF are listed in a table by *Khazragui* and *Hudson* [23.3]. The London School of Economics and Political Science (LSE) maintains a database for research into case studies which were created by the LSE (http://www.lse.ac.uk/reserachAndExpertise/researchImpact/Home.aspx). For example, in the area "society and culture impacts" the database includes case studies on the topics "influencing the cultural diversity of the UK history curriculum", "creating more liveable cities", and "quotas are wrong way to increase female representation on boards". A database with all case studies submitted in the framework of the REF 2014 resides at the following URL: http://impact.ref.ac.uk/CaseStudies/. For example, the area of pure mathematics provides the following case studies: "a benchmark tool for high-performance computing" and "applications of singularity theory and three-dimensional (3-D) modeling in arts and retail".

In a comprehensive study, around 7000 case studies were evaluated, which arose in the framework of the REF 2014 [23.72, 73]. On the one hand the data was submitted to a qualitative analysis, and on the other hand text-mining techniques were used to synthesize the corpus of case studies. The results are summarized as follows [23.72, p. 6]:

> The societal impact of research from UK Higher Education Institutions is considerable, diverse and fascinating. One of the most striking observations from the analysis of the REF case studies was the diverse range of contributions that UK HEIs have made to society ... The relationship between 149 fields of research, 36 UOAs [units of assessments] and 60 impact topics is visualized ... What is evident from this visualization is that multiple fields of research underpin the case studies, leading to multiple types of impact. Overall we identified 3709 unique pathways to impact.

What exactly are the benefits of demonstrating societal impact via case studies in research evaluation? One great advantage is certainly that case studies have no limitations in the representation of societal impact. That is, the method places no restrictions on including any particular kind of societal impact from the start [23.72]. In principle, an institution can report any kind of societal impact. Furthermore, the results of the survey of REF main panel evaluators by *Derrick* [23.74] shows that the "evaluators felt confident that the case studies would prove beneficial in facilitating the evaluation process" [23.74, p. 141]. Case studies offer the possibility of representing the long process of research up to the achievement of societal impact in its complexity [23.31]. This is certainly not possible with any other method. Even if the production of case studies is associated with considerable effort for an institution, the effort seems to be worth it: "It has been reported that a single case study could be worth as much as £720 000 to a university over a 5 year period" [23.3, p. 52].

The effort for the production of case studies may pay off for the institutions, but the large amount of effort is still seen as an important criticism of the method [23.73, p. 150]:

> College London alone wrote 300 case studies that took around 15 person-years of work, and hired four full-time staff members to help, says David Price, the university's vice-provost for research.

An additional criticism of case studies which are produced in the framework of self-evaluation refers to the fact that they generally consist of success stories, which can strictly speaking say almost nothing about the return on total research funding [23.3]. The possibility of writing case studies without a precise specification may lead to a greater variance in the presentation, but it also hinders the production of useful comparisons between the institutions [23.3]. However, only a comparison of institutions makes it clear, which institutions have a better or worse performance than the average (in a country).

Criticism of the case study approach has led in the past to a desire for "a more consistent toolkit of impact metrics that can be more easily compared across and between cases" [23.31, p. 49]. *Atkinson* [23.75] also refers to this point with a concrete example:

> It is arguably harder for the REF to judge and compare quality in this area. There is no guarantee, for example, that a spin-off company that generates 200 jobs and £20 million (US $31 million) in investments will be judged to have more impact than a spin-off that generates 20 jobs and £2 million in investments. Automation is not possible here, but there is room for greater standardization of the dimensions by which impact is assessed and the criteria against which quality is judged.

In the following Sect. 23.3.2, we will describe the need for indicators which we see in the area of societal impact measurement.

23.3.2 The Use of Quantitative Indicators

If we regard the societal impact literature in overview, we see that many studies express dissatisfaction with the currently preferred approach of demonstrating societal impact with case studies (despite the advantages that this approach also offers). Thus, for example, *Ovseiko* et al. [23.55] write:

> The advantage of using quantitative indicators is that they can be standardized and aggregated, allowing universities to use them on a continuous basis to track their impact, compare it with other universities, and recognize the contribution of every faculty member.

According to *Godin* and *Dore* [23.1]:

> We still have, forty years after the first demands for impact indicators, to rely on case studies to quantify, very imperfectly, dimensions other than the economic one [23.1, p. 1].

The National Research Council [23.44] and *Khazragui* and *Hudson* [23.3] feel there is a lack of high-quality metrics, with which the societal (also including the economic) impact of federally funded research can be measured on a national scale. Scientometrics should develop robust indicators:

1. To measure the nature and quality of real world impacts
2. To be able to estimate the longer term benefits of research better
3. To identify the *relevant* research for society [23.12, 23].

The indicators to be developed should be both less labor intensive than case studies, as well as allowing meaningful assessments of research [23.27].

In the development of societal impact indicators, one should take into account that the development of indicators is generally characterized by a laborious iterative process in every area. Indicators must be defined, tested, revised, and validated [23.27]. It must be ensured that these indicators can measure evidence of research impact across different cases in a standardized and consistent manner [23.55]. In the area of societal impact measurement, this standardization will be especially difficult, since the societal impact can vary considerably (Sect. 23.1) [23.47]. An additional difficulty in the development of indicators will be that variants of indicators will be developed, from which it will not initially be clear which of the variants should be favored. The development of the indicators should essentially take into account all disciplines equally [23.76], to avoid a situation which has arisen with academic impact measurement: With the established literature databases (Web of Science, WoS, Clarivate Analytics, and Scopus, Elsevier) an elaborated citation impact measurement is—strictly speaking—only possible in the natural sciences.

Some authors have suggested that a database should be developed for societal impact measurement, which is constructed analogously to the WoS [23.77, p. 604], or a societal impact factor be defined, which resembles the journal impact factor (JIF) in bibliometrics [23.78].

Both suggestions are probably unrealistic (and are also not expedient), since (1) the societal impact measurement is very broad, and it will hardly be possible to measure impact in all areas with only one data source, and (2) the use of the JIF in research evaluation has been sharply criticized (see e. g., the San Francisco declaration on research assessment, DORA, at http://www.ascb.org/dora/). An important reason for the criticism is, for example, that a unit for the determination of the impact of a complete journal should not be used for the impact measurement of the individual contributions published in it. The overall impact does not correspond to the impact of the individual contributions in most cases. Societal impact measurement should not repeat the mistake in academic impact measurement that an indicator is developed or used which is subject to sharp criticism.

In Sect. 23.4, we provide an overview of the data sources and metrics from the area of altmetrics, which could be attractive for societal impact measurement. Altmetrics is a new topic which has only been investigated for a few years in scientometrics. The proponents of altmetrics often produce the impression that the first credible measurement of societal impact began with altmetrics. But, as the literature shows, indicators for the measurement of societal benefits of research have been suggested and applied for many years. *Molas-Gallart* et al. [23.27] suggest specifying the measurement of the societal impact of an individual department or university by the number of joint publications (academic publications) between university and industrial firms. These joint publications can be understood as productive interactions which are expected to produce considerable impact. *van der Meulen* and *Rip* [23.45] further expand the measurement of collaborations via publications with the following areas in which there could be direct relations with societal actors: collaborative projects, contract research, membership of advisory boards and committees.

To enable a reliable, fair, and valid societal impact measurement, the following should be considered in indicator developments:

- The indicators should in fact only measure what they claim to measure [23.17]. Since societal impact measurement is often very broadly construed, a restriction to the relevant areas is desirable [23.55]:

  > The issue of validity relates to the degree of certainty that the proposed indicators measure what they claim to measure, i. e., research impact. Without applying precise measures of impact and criteria for the attribution of impact to specific research activities across all universities, any impact assessment will be inconsistent and, thus, unreliable.

- Societal impact measurement is significantly more complex than academic impact measurement, which is primarily a matter of quality. According to *Smith* [23.76, p. 22]:

  > When research has to deal with society, however, a wide range of issues from ethics and safety to economics, legal issues and politics also come into play.

- The results of societal impact measurement should preferably be compared with a control situation, e. g., without the evaluated research activity. This comparison should reveal the effect an individual research activity really has [23.11].
- The indicators should be *socially robust*. *Barré* [23.17] considers that they can be understood in different contexts by stakeholders. These stakeholders could include [23.76]:

  1. Policy-makers at the intermediary or government level
  2. Professional users in industry and societal organizations (developing products and services)
  3. End users (the public at large or individual target groups).

- The use of indicators in research evaluation can lead to an adjustment of the behavior of the evaluated scientists [23.15, 23, 79]. This adjustment may or may not be intentional. Thus, for example, there is the danger in the development of indicators for societal impact measurement that indicators are suggested that are easy to measure but which can say little about the societal impact of research. The use of these indicators could result in the adjustment of scientists to these indicators which leads to a worsening of the societal benefit of research [23.23].
- With societal impact measurement it should be taken into account that different disciplines can expect different levels of societal impact. This is mainly because some disciplines are more productive and make a far greater contribution to society than others [23.40]. The impact to be expected is therefore context-dependent and should be measured in a context-dependent way [23.45]. The impact model may work in some disciplines (e. g., medicine or engineering), but many disciplines have other forms of impact through practices. One can imagine, for example, semantic maps (based on coword analysis) to provide richer means for societal impact assessment in these cases.
- *Molas-Gallart* et al. [23.27] regard it as necessary that the indicators for societal impact measurement be normalized for the size of a research unit: “Normalising by staff numbers provides a uni-

form procedure for indicators and therefore reduces the complexity of the scoring system" [23.27, p. 53].

- The indicator set for societal impact measurement should not be too extensive [23.80]. Since societal impact measurement is a wider field than academic impact measurement, societal impact measurement will certainly require more indicators than academic impact measurement. However, the set should still remain manageable.

23.4 Altmetrics

With the measurement of citations in patents and in clinical guidelines, two promising altmetrics for societal impact measurement have already been suggested and applied [23.81]. Social media metrics could also have potential for societal impact measurement. Even if some social media data (e. g., Mendeley reader counts) carry similar information as citation counts [23.82] and are more associated with academic than with societal impact, most social media data other than online reference manager counts very likely measures other kinds of impact than academic impact [23.83]. Thus, references to scientific papers in policy-related documents may be applied to measure the impact of papers on policy-related areas [23.84, 85].

In the following sub-sections, we will discuss social media metrics, citations in patents, citations in clinical guidelines, and references to scientific papers in policy-related documents in more detail for their potential use for measurement of societal impact. These four possibilities for societal impact measurement are mainly interesting for research evaluation because there are parallels with the measurement of academic impact by means of citations. The similarity of the data allows established methods and procedures of evaluation to be transferred from bibliometrics to altmetrics.

23.4.1 Social Media Metrics

Social media metrics constitute the core of altmetrics. *Robinson-Garcia* et al. [23.86] found that Twitter and Mendeley are the most important social media metrics when judging by the coverage of scientific papers in various sources tracked by Altmetric. Altmetric is a digital science company based in London that tracks and analyzes the online activity around scholarly research outputs (https://www.altmetric.com). Some social media metrics (including Twitter and Facebook counts) have already been included in the *Snowball Metrics Recipe Book* [23.87]. In this book, some universities (in cooperation with Elsevier) try to standardize their way of measuring institutional output and impact.

Some promising empirical results have been published which reveal the potential of altmetrics for societal impact measurements. *Bornmann* and *Haunschild* [23.88] combined expert-based evaluations about published scientific papers (post publication peer review as performed by F1000Prime) with altmetrics data. They found that in the case of a well-written article that provides a good overview of a topic, it tends to be better received by people outside research—measured by altmetrics. Similarly, *Bornmann* [23.69] reports that papers tailored for a readership outside the area of research lead to measurable societal impact. *Bornmann* [23.89] has shown that counts of tweets and Facebook posts might be useful for the measurement of societal impact. A meta-analysis of studies correlating citations with different altmetrics by *Bornmann* [23.90, p. 1140] points out:

> That the more a social media community is dominated by people focusing on research, the higher the correlation between the corresponding altmetric and traditional citations is.

A low pooled correlation between traditional citations and microblogging was found which indicates potential use for societal impact measurement. This conclusion was substantiated in the meta-analysis by *Erdt* et al. [23.91].

As normalization of altmetrics data is important when said data is to be used for impact measurement, *Bornmann* and *Haunschild* [23.92] proposed a methodology to normalize Twitter counts based on percentiles and introduced the Twitter index. Other normalization procedures were introduced for Mendeley readership data [23.93–96]. These normalization methods are based on the calculation of the average number of reader counts per scientific field and publication year. The ratio of the raw number of reader counts and the average number of reader counts of the same scientific field and publication year yields a normalized reader score. However, normalization methods based on averages are problematic when sparse altmetrics data are considered [23.97]. Thus, *Thelwall* [23.98], *Haunschild* and *Bornmann* [23.99], and *Bornmann* and *Haunschild* [23.100] proposed metrics based on the proportion of mentioned papers. In contrast to average-based methods, the methods based on mentioned proportions might be able to handle sparse altmetrics data, such as mentions in Wikipedia pages.

As the part of society represented by many social media metrics (especially tweets and Facebook posts) could not yet be defined accurately, their use for societal impact measurement should be questioned [23.101]. Furthermore, adherence to scientific citation habits cannot be expected for most altmetrics sources. Gaming metrics is much easier for social media metrics than for citation-based metrics in scientific documents [23.102]. Also, social media metrics are easier to manipulate using automated accounts and bots than citation-based metrics. This problem has been studied by *Haustein* et al. [23.103] for Twitter. Their results show that automated Twitter accounts create a considerable amount of tweets mentioning scientific papers. They also conclude that automated Twitter accounts and bots have critical implications for the use of raw Twitter counts in research evaluation.

Overall, certain social media metrics seem to be suitable sources for societal impact measurement. A promising approach to societal impact measurement is broad (many different sources) but target-oriented (using a specific group of the society). For example, Mendeley users can be assigned to certain groups of society based on their (academic) status groups [23.94]. Thus, it is possible to measure the impact of research on certain target groups in society (e.g., students). These measurements can be made time- and field-normalized. However, more research is necessary to identify standard approaches based on social media metrics which may be useful for measurement of societal impact of research papers [23.81, 102, 104, 105].

23.4.2 Citations in Patents

Patents are official documents that describe inventions (e.g., solutions to a specific technological problem, a new product, or an innovative process). Patent counts themselves have already been investigated as an indicator of societal impact: "A classical 'proxy' for measuring the technological orientation of a public research institution or a university, is to build an indicator of its patenting activity" [23.17, p. 128].

Similar to publications, patents also contain references. As most nonpatent references in patents are journal references [23.106], societal (economic) benefit from research might be demonstrable by using methods from traditional bibliometrics (citations from journal publications to journal publications). Thus, citations from patents to scientific publications may be used as indicators for societal impact measurement. Specifically, citations in patents may be used to assess the contribution of publicly funded research to innovations in industry [23.55]. *Kousha* and *Thelwall* [23.107] compared citations from patents to scientific publications with citations from scientific publications to scientific publications for 16 different research fields and found a low but positive correlation. Apparently, patent citations provide a different perspective of impact measurement than citations from scientific publications.

In a classical study, *Narin* et al. [23.108] examined the front pages of 400 000 US patents issued in 1987/1988 and 1993/1994, and traced the 430 226 nonpatent references contained in these patents, of which 242 000 were judged to be scientific references and 175 000 were to papers published in the (back then) 4000 journals covered by the science citation index. They found that many of the scientific references were funded by research organizations. However, not only universities but also top industrial laboratories (IBM, General Electric, Motorola, Xerox, and Hewlett Packard) published papers that were referenced in a substantial proportion of the analyzed patents. They also noted three other interesting points:

1. The research papers cited in patents originate predominantly from prestigious universities and are published in prestigious journals.
2. Patents with applicants from a certain country cite publications with authors from the same country more often than from other countries (about two to four times more often). Thus, they found a significant nationality bias.
3. The number of citations to US authored papers has tripled within six years (1987/1988 citing 1975–1985 versus 1993/1994 citing 1981–1991).

In a sophisticated, more recent patent study, *Chang* and *Breitzman* [23.109] focus on patent to patent citations. They provide a methodology to distinguish between hot patents and next-generation patents. They apply a cocitation analysis to cluster patents and rank the resulting clusters by the likelihood of the clusters to contain emerging technologies. Such analyses make it possible to measure the socioeconomic impact of science target-oriented on emerging technologies by determination of references to scientific publications in specific patents of clusters with a high likelihood of emerging technologies.

Analysis of patents and patent citations are not without limitations: (1) not all inventions are patented and (2) patenting habits vary across technological fields [23.106, 109]. The latter limitation may be overcome by normalization procedures. As it has become standard in bibliometrics for journal to journal citations, also analysis of patent to journal citations (as well as patent to patent citations) can employ normalization methods to account for differences in scientific discipline (and/or patent class) and publication years [23.70, 110].

Patent counts and (normalized) citations from patents to scientific publications seem to offer a promising route for societal impact measurement into the economic sector of science. SciVal, a commercial research evaluation tool by Elsevier, uses patent-based metrics [23.111, p. 2]:

> SciVal looks at the citations of scholarly output in patents and provides links to both the citing patents and cited Scopus articles. This helps showcase connections between science and industry as well as the knowledge flows.

23.4.3 Citations in Clinical Guidelines

Clinical guidelines are documents which aim to guide decisions and criteria regarding diagnosis, management, and treatment in healthcare. These documents have been in use for the entire history of medicine in one form or another. However, in contrast to earlier forms, which were often based on tradition or authority, modern clinical guidelines are based on an examination of current evidence within the paradigm of evidence-based medicine. They usually include summarized consensus statements on best practices in healthcare. Their formulations bring pieces of important and influential research together. It is an indication for funders that research they have supported is likely to be influencing medical policy and clinical practice, if the funded research is referenced as part of the evidence supporting a national and/or international clinical guideline.

Grimshaw and *Russell* [23.112] have analyzed 59 studies that evaluated clinical guidelines. They concluded that clinical guidelines do improve medical practice. It has been recommended that medical experts come together periodically to review existing evidence and formulate official guidelines for preventive care, diagnosis, and treatment of diseases. Such practices not only help to identify and disseminate best practices for treating patients but also increase the potential value of clinical guidelines for societal impact measurements. Also existing clinical guidelines should be reviewed and if necessary updated on a regular basis [23.113].

Andersen [23.114] analyzed 80 clinical guidelines which were rated regarding their "rigor of development". He extracted 5970 scientific publications indexed in the WoS from the reference lists of the clinical guidelines. Computing two different kinds of normalized citation scores (one based on averages and one based on percentiles), he found a significantly positive correlation between both normalized citation scores and the expert ratings regarding the rigor of development of the clinical guidelines. This indicates that the quality of clinical guidelines correlates with the quality of the scientific publications they are based on. This is in agreement with *Thelwall* and *Maflahi* [23.115] who found that papers in clinical guidelines are much more likely to be cited more often than comparable papers not referenced in clinical guidelines.

In a pilot study on citations in clinical guidelines, *Grant* [23.14] analyzed clinical guidelines with regard to outcomes of biomedical research on healthcare. He found that out of a total of 284 publications, which were referenced in the reference lists of clinical guidelines, 273 (or 96%) were papers in research journals. 235 (or 86%) of these papers were found in library catalogues and 154 (or 65%) had a funding acknowledgement. Therefore, reference lists of clinical guidelines and the funding attributions in the referenced publications are potentially an attractive resource to support research evaluation regarding societal impact [23.116]. Usage of clinical guidelines means demonstrating an improvement in healthcare for biomedical agencies. However, there are potential problems as there exists a complex relationship between research and its incorporation into new treatment guidelines. The main conclusion from his pilot study is that it is possible to apply conventional bibliometric techniques to trace the transfer of knowledge from research funding into clinical practice.

Lewison and *Sullivan* [23.117] studied UK clinical guidelines relevant to cancer (43 documents) as well as 3217 references cited therein. These references were mainly published in high-impact journals, but this might be a circular effect [23.117, p. 1948]:

> It appears that if researchers want their work, particularly clinical trials, to be part of the evidence base for clinical guidelines, then it is desirable for them to publish in highly cited journals.

Lewison and *Sullivan* [23.117] used the cited reference data to investigate the contribution of countries to clinical guidelines relevant to cancer. They found that UK papers were referenced three times as frequently as one would expect from their occurrence in the world oncology literature. Thus, a similar effect as found by *Narin* et al. [23.108] for patents might have been observed by *Lewison* and *Sullivan* [23.117] for clinical guidelines.

Thelwall and *Maflahi* [23.115] found that scientific publications mentioned in clinical guidelines are also often highly cited in the academic literature, especially in the case of publications older than three years. They see a potential benefit from the use of clinical guideline citations especially for recent publications. According to *Jones* and *Hanney* [23.118], societal impact studies based on citations from clinical guidelines usually only focus "on the direct influence from the paper cited, and it is argued that impacts usually arise from one or more streams of research or from a variety of papers" [23.118, pp. 976/977]. They investigated the indirect societal impact of biomedical research by studying several citation generations from key research articles and clinical guidelines.

Assessing the value of medical funding is of increasing importance [23.115]. Besides analysis of citations from clinical guidelines, also mentions of healthcare funders are used in impact measurement. In the latter cases, the societal impact of research funded by healthcare organizations is measured: Funders of medical research have made efforts to enhance the understanding of the impact of their funded research and to provide evidence of the value of investments in particular areas of research. Research evaluation around impact on policy and practice of medical treatments represents one of the most challenging areas.

Medical research funders have already started to track clinical guidelines as potential indicators of societal impact. *Kryl* et al. [23.116] analyzed authorship and funding attribution of scientific publications which are part of the reference lists of two clinical guidelines from the National Institute of Health and Clinical Excellence (NICE). They found that about a third of the scientific publications are authored by at least one scientist in the UK and about half of these publications from the UK listed a diverse set of funders in their acknowledgment sections. They conclude that reference lists of clinical guidelines offer great potential for the quality assessment of scientific publications. In principle, a presence in clinical guideline reference lists could serve as evidence that a research funder's money has been useful for society in ways other than scientific knowledge building. They recognize tracking and harnessing the relevant information in a reliable way as one of the main challenges.

Bunn et al. [23.119] have studied societal and scientific impact of Cochrane Reviews. Cochrane Reviews contain scientific publications as well as clinical guidelines in their reference lists. Cochrane Review Groups (CRGs) gather and summarize medical evidence from research. The aim is to provide help to make informed choices about medical treatments. The findings of CRGs are available in the Cochrane Database of Systematic Reviews (http://www.cochranelibrary.com/). *Bunn* et al. [23.119] found that Cochrane Reviews are used to inform healthcare policy-makers and are helpful during development and improvement of clinical guidelines. Among the benefits that Cochrane Reviews appeared to have had, *Bunn* et al. [23.119] mentioned safer or more appropriate use of medication or other healthcare technologies, or the identification of new effective drugs or treatments. It is unclear, however, whether or not these changes were directly a result of the Cochrane Reviews and not the result of subsequent clinical guidelines. In any case, the reference lists of Cochrane Reviews seem to be another promising source for societal impact measurement.

Although references in clinical guidelines to scientific documents are useful in societal impact measurement, there are limitations, as clinical guidelines do not always contain references. Furthermore, the relevance and value of the listed references is not explicitly indicated [23.116]. However, this is not different to citations in scientific publications or mentions in various altmetrics sources. Still, the value of clinical guidelines for societal impact measurement could be increased if research funders and policy-makers would advocate for references to be systematically added to and labeled in clinical guidelines. Other limitations are likely to remain:

1. Citation of review articles rather than the original studies
2. Biases in the selection committees of clinical guidelines
3. Citing follow-up studies rather than the original research articles
4. Standardized procedures for updating clinical guidelines are not implemented [23.115]. Clinical guidelines may lose their clinical relevance as they age and newer research emerges. Outdated clinical guidelines should be retracted but often are not.
5. An analysis of scientific papers stating support from healthcare funders may be incomplete and papers might acknowledge funding although the funding did not provide any substantive contribution for the paper and subsequently for the improvement of healthcare.

Despite these limitations, clinical guidelines seem to be an attractive resource for societal impact measurement into the healthcare sector of science.

23.4.4 References in Policy-Related Documents

Ritter and *Lancaster* [23.120] argue that assessment of the extent to which "research influences and impacts policy decision making needs to go beyond bibliometric analysis of academic citations" [23.79, p. 30]. They recommend a systematic analysis of policy documents.

Policy-related documents (policy documents) are published by governmental agencies as well as by nongovernmental organizations (NGOs). The types of NGOs publishing policy documents varies from large international organizations (such as the World Health

Organization, see http://www.who.int) to small local organizations (such as the Economic Commission for Latin America and the Caribbean, see http://www.cepal.org/en) and special interest groups (such as think tanks and lobbying groups). Policy documents authored by NGOs are mainly directed to leaders and decision-makers in politics in order to shape political decisions, while policy documents published by governments are mainly directed to the general public to support and explain decision-making in politics. Modern governments exhibit scientific service facilities and NGOs employ (depending on their organization's size) scientific staff members to help draft policy documents. Irrespective of the author of a policy document, such documents contain three crucial sections:

1. A current problem is presented
2. A particular course of action is recommended
3. Supporting evidence for the recommendation is cited. As a variety of different people work on a single policy document, one should not expect scientific citation habits in policy documents.

Analysis of reference lists of policy documents on a large scale were made possible by aggregators of altmetrics. Altmetric started to provide mentions of scholarly papers in policy documents in 2014. *Liu* [23.121] writes:

> As you might have already learned from our June press release announcing the launch of Altmetric for institutions, we recently started tracking some highly impactful new sources of attention: policy and guidance documents.

Plum Analytics started to provide similar information in December 2016 [23.122]. *Haunschild* and *Bornmann* [23.85] analyzed how many and which proportion of papers are mentioned in policy documents tracked by Altmetric. They found that less than 0.5% of the papers published in different subject categories are mentioned at least once in policy documents using a large interdisciplinary database (WoS). *Bornmann* et al. [23.84] studied more specifically how many and which proportion of papers from climate change literature were mentioned in policy documents. As climate change research is a rather politicized and important topic, many mentions in policy documents could be expected. However, only 1.2% out of 191 276 papers on climate change were found to have at least one policy mention.

Although an indicator based on papers mentioned in policy documents seems to be a promising indicator for societal impact measurement, several problems exist currently:

1. Mainly, international, English documents are tracked by data providers. Thus, more non-English policy sources should be tracked.
2. Policy documents are often published in different languages as different documents but with the same content. Disambiguation of such duplicates is problematic.
3. It is unknown where a particular paper has been mentioned in a policy document or policy source website. For example, the mention could be in a publication list in a CV instead of a policy document.

More studies should investigate coverage of policy documents and their mentions of scholarly material and solutions to the aforementioned problems should be found before mentions in policy documents are used for societal impact measurement.

23.5 Discussion

As a part of society, science is also affected by the overall development towards an audit society. It is no longer taken for granted that science makes an important contribution to society; this contribution must be demonstrated or proven. Here, it is chiefly a matter of the production of direct societal and economic utility and impact. For science, this means that research evaluation must go beyond the established instruments of peer review and bibliometrics or that the established instruments must be adapted to the new conditions. One possible adaptation could be to include people from areas outside science in peer review panels. As the results of *Ballabeni* et al. [23.123] show, the societal relevance of research is nothing new for many scientists and plays an important role in the motivation for pursuing research. The authors surveyed more than 300 scientists at Harvard Medical School and affiliated institutes [23.123]:

> The majority of the scientists who participated in the survey indicated that the most important goal of publicly funded basic 'biomedical' research is the production of health benefits to society (86%) ... and that the desire to effectively benefit society

is an important or very important motivation for most of them (87%).

The measurement of societal impact can be regarded as a "Kuhnian revolution for research evaluation criteria" [23.74, p. 137] see also [23.124, 125]. Societal impact measurement is a matter of the measurable influence (effect) of research which can be detected outside of research in a particular sector of society. Besides the accountability in an audit society, this chapter mentions additional reasons for the desire for societal impact measurement: science should be more strongly oriented towards the application or use of its results. This is mainly in the hope of strengthening the economy of a country. Furthermore, research projects should be promoted when these research projects help to tackle specific societal problems.

However, research evaluation and funding decisions should always be carried out in a contextualized manner. Most fundamental research has no measurable societal impact during useful evaluation periods, but most fundamental research projects do have significant societal impact over the long run (e. g., climate change research and space sciences). Since research happens in very disparate areas and—besides fundamental research—there is also application-oriented research, societal impact measurement should take account of the breadth of the possible usefulness of research. However, societal impact measurement is afflicted with a number of problems which we discussed in Sect. 23.1.3. These include:

1. The difficulty with the attribution of societal impact to a particular piece of research
2. The possibility that the impact of research on society can be both positive and negative at the same time
3. The often long time lag between the performance of the research and its (measurable) effect on society. These problems should always be taken into account for societal impact measurement.

In Sect. 23.2 we have given an overview of societal impact measurement in various areas (with funding bodies, in national evaluation exercises as well as through multidimensional frameworks and productive interactions). As we have set out in Sect. 23.3, case studies are currently the method favored for societal impact measurement. Case studies offer a number of advantages, such as the possibility of acquainting an audience which does not consist of specialists with complex issues. In addition, case studies can be used in all disciplines—without restriction. Case studies, however, have some serious disadvantages: their results cannot be generalized nor are they suitable for the comparison of research units. They are also relatively expensive to perform. Therefore, it is desirable (in research evaluation) that quantitative indicators for societal impact measurement be developed, the results of which can both be generalized and permit a comparison between different research units. When these indicators are available, they should not replace the case studies approach completely, but complement it in a mixed-methods approach [23.64].

Section 23.4 therefore discusses which altmetrics have been suggested or are already used for the measurement of societal impact. Sections 23.4.2 and 23.4.3 describe studies which show that citations in patents and citations in clinical guidelines can already be used in certain areas for societal impact measurement. Clinical guidelines in particular extract information from research literature quickly [23.82]. A great advantage of citations in patents and citations in clinical guidelines is that they contain citation data which can be evaluated with the established methods of bibliometrics. Section 23.4.4 discusses the potential use of mentions of papers in policy-related documents for societal impact measurement. The documents also contain reference lists which can be analyzed with methods from bibliometrics. However, too few studies have been carried out thus far so that it is an open question as to whether policy-related documents will be useful in societal impact measurement although they look rather promising at this point.

Social media metrics—as presented in Sect. 23.4.1—are also seen as a possibility for measuring the societal impact of research. It is however not yet clear whether social media metrics can really satisfy this requirement. We assume that only a subset of social media metrics can actually be used for societal impact measurement. Which metrics these will be in practice, will be shown by scientometric research. In addition, we do not regard it as sensible to use raw counts for impact measurement or to combine the raw counts of various altmetrics sources into composite indicators, but rather to undertake a field- and time-normalized impact measurement, which is target-oriented (and focused on such groups in society as students or others).

References

23.1 B. Godin, C. Dore: Measuring the Impacts of Science: Beyond the Economic Dimension, HIST Lecture Helsinki Institute for Science and Technology Studies, Helsinki, Finland, http://www.csiic.ca/PDF/Godin_Dore_Impacts.pdf (2005)
23.2 L. Bornmann: Scientific peer review, Annu. Rev. Inf. Sci. Technol. **45**, 199–245 (2011)
23.3 H. Khazragui, J. Hudson: Measuring the benefits of university research: Impact and the REF in the UK, Res. Eval. **24**(1), 51–62 (2015)
23.4 S.P.L. de Jong, P. van Arensbergen, F. Daemen, B. van der Meulen, P. van den Besselaar: Evaluation of research in context: An approach and two cases, Res. Eval. **20**(1), 61–72 (2011)
23.5 S. Mostert, S. Ellenbroek, I. Meijer, G. van Ark, E. Klasen: Societal output and use of research performed by health research groups, Health Res. Policy Syst. **8**(1), 30 (2010)
23.6 M. Power: *The Audit Society: Rituals of Verification* (Oxford Univ. Press, Oxford 1999)
23.7 G. Gruening: Origin and theoretical basis of new public management, Int. Public Manag. J. **4**(1), 1–25 (2001)
23.8 J.C. Petit: Why do we need fundamental research?, Euro. Rev. **12**(2), 191–207 (2004)
23.9 S. Hanney, T. Packwood, M. Buxton: Evaluating the benefits from health research and development centres: A categorization, a model and examples of application, Evaluation **6**(2), 137–160 (2000)
23.10 M. Jubb: The scholarly ecosystem. In: *Academic Professional Publishing*, ed. by R. Campbell, E. Pentz, I. Borthwick (Chandos, Oxford 2012) pp. 53–77
23.11 A.J. Milat, A.E. Bauman, S. Redman: A narrative review of research impact assessment models and methods, Health Res. Policy Syst. (2015), https://doi.org/10.1186/s12961-015-0003-1
23.12 G. Cohen, J. Schroeder, R. Newson, L. King, L. Rychetnik, A.J. Milat, S. Chapman: Does health intervention research have real world policy and practice impacts: Testing a new impact assessment tool, Health Res. Policy Syst. **13**, 12 (2015)
23.13 C. Donovan: The Australian research quality framework: A live experiment in capturing the social, economic, environmental, and cultural returns of publicly funded research, New Dir. Eval. **2008**(118), 47–60 (2008)
23.14 J. Grant: Evaluating the outcomes of biomedical research on healthcare, Res. Eval. **8**(1), 33–38 (1999)
23.15 L. Bornmann: Mimicry in science?, Scientometrics **86**(1), 173–177 (2011)
23.16 J.B. Spaapen, H. Dijstelbloem, F. Wamelink: *Evaluating Research in Context: A Method for Comprehensive Assessment* (Consultative Committee of Sector Councils for Research and Development, The Hague 2007)
23.17 R. Barré: S and T indicators for policy making in a changing science-society relationship. In: *Handbook of Quantitative Science and Technology Research*, ed. by H. Moed, W. Glänzel, U. Schmoch (Springer, Dordrecht 2005) pp. 115–131
23.18 K. Lähteenmäki-Smith, K. Hyytinen, P. Kutinlahti, J. Konttinen: *Research with an Impact Evaluation Practises in Public Research Organisations* (VTT Technical Research Centre of Finland, Kemistintie 2006)
23.19 P. Dahler-Larsen: *The Evaluation Society* (Stanford Univ. Press, Stanford 2011)
23.20 G.M. Lamm: Innovation Works™. A case study of an integrated pan-European technology transfer model, B.I.F. Futura **21**(2), 86–90 (2006)
23.21 P. Nightingale, A. Scott: Peer review and the relevance gap: Ten suggestions for policy-makers, Sci. Public Policy **34**(8), 543–553 (2007)
23.22 J. Ziman: Why must scientists become more ethically sensitive than they used to be?, Science **282**(5395), 1813–1814 (1998)
23.23 E. Ernø-Kjølhede, F. Hansson: Measuring research performance during a changing relationship between science and society, Res. Eval. **20**(2), 131–143 (2011)
23.24 J. Cohen: A hard look at global health measures, Science **345**(6202), 1260–1265 (2014), https://doi.org/10.1126/science.345.6202.1260
23.25 Toronto Region Research Alliance: *The Race for Global Leadership in Innovation: An Analysis of National R and D Strategies* (TRRA, Toronto 2001)
23.26 M. Cleary, N. Siegfried, D. Jackson, G.E. Hunt: Making a difference with research: Measuring the impact of mental health research, Int. J. Ment. Health Nurs. **22**(2), 103–105 (2013)
23.27 J. Molas-Gallart, A. Salter, P. Patel, A. Scott, X. Duran: *Measuring Third Stream Activities. Final Report to the Russell Group of Universities* (SPRU, Brighton 2002)
23.28 I. Peters, A. Jobmann, A. Eppelin, C.P. Hoffmann, S. Künne: Altmetrics for Large, Multidisciplinary Research Groups: A Case Study of the Leibniz Association, https://www.alexandria.unisg.ch/Publikationen/nach-neuen-Publikationen/231796 (2014)
23.29 ERiC: *Evaluating the Societal Relevance of Academic Research: A Guide* (Delft Univ. Technology, Delft 2010)
23.30 B. Göransson, R. Maharajh, U. Schmoch: New activities of universities in transfer and extension: Multiple requirements and manifold solutions, Sci. Public Policy **36**(2), 157–164 (2009)
23.31 J. Wilsdon, L. Allen, E. Belfiore, P. Campbell, S. Curry, S. Hill, B. Johnson: *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management* (HEFCE, Bristol 2015)
23.32 A. Dance: Impact: Pack a punch, Nature **502**, 397–398 (2013)
23.33 G.N. Samuel, G.E. Derrick: Societal impact evaluation: Exploring evaluator perceptions of the

characterization of impact under the REF2014, Res. Eval. **24**(3), 229–241 (2015)
23.34 J. Grimson: Measuring research impact: Not everything that can be counted counts, and not everything that counts can be counted. In: *Bibliometrics: Use and Abuse in the Review of Research Performance*, ed. by W. Blockmans, L. Engwall, D. Weaire (Portland, London 2014) pp. 29–41
23.35 T. Penfield, M.J. Baker, R. Scoble, M.C. Wykes: Assessment, evaluations, and definitions of research impact: A review, Res. Eval. **23**(1), 21–32 (2014), https://doi.org/10.1093/reseval/rvt021
23.36 Advisory Group: *Research Quality Framework: Assessing the Quality and Impact of Research in Australia*. *Research Impact* (Department of Education, Canberra 2006)
23.37 Committee for Scientific and Technology Policy: *Assessing the Impact of State Interventions in Research—Techniques, Issues and Solutions* (Directorate for Science, Technology and Innovation, Brussels 2014)
23.38 R. Ruegg, I. Feller: *A Toolkit for Evaluating Public R and D Investment: Models, Methods, and Findings from ATP's First Decade* (National Institute of Standards and Technology, Gaithersburg 2003)
23.39 B.R. Martin: Assessing the impact of basic research on society and the economy. In: *Rethink. Impact Basic Res. Soc. Economy, Vienna* (2007)
23.40 SISOP: Conceptualizing the Social Impact of Science, http://sisobproject.wordpress.com/2011/07/20/conceptualizing-the-social-impact-of-science-2/ (2011)
23.41 A.J. Salter, B.R. Martin: The economic benefits of publicly funded basic research: A critical review, Res. Policy **30**(3), 509–532 (2001)
23.42 Higher Education Funding Council for England: *Research Excellence Framework. Second Consultation on the Assessment and Funding of Research* (HEFCE, Bristol 2009)
23.43 L. Langfeldt, L. Scordato: *Assessing the Broader Impacts of Research: A Review of Methods and Practices* (NIFU, Oslo 2015)
23.44 National Research Council: *Furthering America's Research Enterprise* (The National Academies, Washington 2014)
23.45 B. van der Meulen, A. Rip: Evaluation of societal quality of public sector research in the Netherlands, Res. Eval. **9**(1), 11–25 (2000)
23.46 Health Economics Research Group, O.o.H.E., RAND Europe: *Medical research: What's it worth? estimating the economic benefits from medical research in the UK* (Evaluation Forum, London 2008)
23.47 B. Gregersen, L.T. Linde, J.G. Rasmussen: Linking between Danish universities and society, Sci. Public Policy **36**(2), 151–156 (2009)
23.48 A. Oancea: Buzzwords and values: The prominence of "impact" in UK research policy and governance, Res. Trends (33), 6–8 (2013)
23.49 SIAMPI: SIAMPI Workshop—Brussels, http://www.siampi.eu/Content/SIAMPI/Report%20SIAMPI%20workshop.pdf (2010)
23.50 L. Bornmann: Measuring impact in research evaluations: A thorough discussion of methods for, effects of and problems with impact measurements, High. Educ. **73**(5), 775–787 (2017)
23.51 J. Ziman: *Real Science. What it is, and What it Means* (Cambridge Univ. Press, Cambridge 2000)
23.52 G. Bechmann, M. Decker, U. Fiedeler, B.J. Krings: Technology assessment in a complex world, Int. J. Foresight Innov. Policy **3**(1), 6–27 (2007)
23.53 B.R. Martin, J. Irvine: Assessing basic research—Some partial indicators of scientific progress in radio astronomy, Res. Policy **12**(2), 61–90 (1983)
23.54 L. Rymer: *Measuring the Impact of Research—The Context for Metric Development* (The Group of Eight, Turner 2011)
23.55 P.V. Ovseiko, A. Oancea, A.M. Buchan: Assessing research impact in academic clinical medicine: A study using research excellence framework pilot impact indicators, BMC Health Serv. Res. (2012), https://doi.org/10.1186/1472-6963-12-478
23.56 J.B. Holbrook, S. Hrotic: Blue skies, impacts, and peer review, RT A J. Res. Policy Eval. (2013), https://doi.org/10.13130/2282-5398/2914
23.57 G.E. Derrick, I. Meijer, E. van Wijk: Unwrapping "impact" for evaluation: A co-word analysis of the UK REF2014 policy documents using VOSviewer. In: *Proc. Sci. Technol. Indic. Conf. 2014, Leiden "Context Counts: Pathways to Master Big and Little Data"*, ed. by P. Wouters (Univ. Leiden, Leiden 2014) pp. 145–154
23.58 M. Knott: Academic Publications to Become Less Important when Funding University Research, The Sydney Morning Herald (15.11.2015), http://www.smh.com.au/federal-politics/political-news/academic-publications-to-become-less-important-when-funding-university-research-20151112-gkxkgl
23.59 L. Bornmann: Measuring the societal impact of research, EMBO Rep. **13**(8), 673–676 (2012)
23.60 L. Bornmann: What is societal impact of research and how can it be assessed? A literature survey, J. Am. Soc. Inf. Sci. Technol. **64**(2), 217–233 (2013)
23.61 M. Buxton, S. Hanney: How can payback from health services research be assessed?, J. Health Serv. Res. Policy **1**(1), 35–43 (1996)
23.62 C. Donovan, S. Hanney: The 'Payback Framework' explained, Res. Eval. **20**(3), 181–183 (2011)
23.63 C. Donovan: State of the art in assessing research impact: Introduction to a special issue, Res. Eval. **20**(3), 175–179 (2011)
23.64 R. Miettinen, J. Tuunainen, T. Esko: Epistemological, artefactual and interactional–institutional foundations of social impact of academic research, Minerva **53**(3), 257–277 (2015)
23.65 S. de Jong, K. Barker, D. Cox, T. Sveinsdottir, P. Van den Besselaar: Understanding societal impact through productive interactions: ICT research as a case, Res. Eval. **23**(2), 89–192 (2014)
23.66 SIAMPI: Final Report on Social Impacts of Research, http://www.siampi.eu/Content/SIAMPI_Final%20report.pdf. Last Access: 11/1/2018 (2011)

23.67 C.H. Llewellyn Smith: What's the Use of Basic Science?, http://wwwnew.jinr.ru/section.asp?sd_id=94 (1997)
23.68 L. Bornmann, W. Marx: How should the societal impact of research be generated and measured? A proposal for a simple and practicable approach to allow interdisciplinary comparisons, Scientometrics **98**(1), 211–219 (2014)
23.69 L. Bornmann: Validity of altmetrics data for measuring societal impact: A study using data from Altmetric and F1000Prime, J. Informetr. **8**(4), 935–950 (2014)
23.70 J.A. Wallin: Bibliometric methods: Pitfalls and possibilities, Basic Clin. Pharmacol. Toxicol. **97**(5), 261–275 (2005)
23.71 C.M. Pålsson, B. Göransson, C. Brundenius: Vitalizing the Swedish university system: Implementation of the 'third mission', Sci. Public Policy **36**(2), 145–150 (2009)
23.72 L. King's College, D. Science: *The Nature, Scale and Beneficiaries of Research Impact: An Initial Analysis of Research Excellence Framework (REF) 2014 Impact Case Studies* (King's College London, London 2015)
23.73 R. Van Noorden: Seven thousand stories capture impact of science, Nature **518**(7538), 150–151 (2015)
23.74 G. Derrick: Intentions and strategies for evaluating the societal impact of research: Insights from REF2014 evaluators. In: *Proc. Sci. Technol. Indic. Conf. 2014, Leiden "Context Counts: Pathways to Master Big and Little Data"*, ed. by P. Wouters (Univ. Leiden, Leiden 2014) pp. 136–144
23.75 P.M. Atkinson: Assess the real cost of research assessment, Nature **516**(7530), 145–145 (2014)
23.76 R. Smith: Measuring the social impact of research, BMJ **323**(7312), 528 (2001)
23.77 J.M. Bensing, W.M.C.M. Caris-Verhallen, J. Dekker, D.M.J. Delnoij, P.P. Groenewegen: Doing the right thing and doing it right: Toward a framework for assessing the policy relevance of health services research, Int. J. Technol. Assess. Health Care **19**(04), 604–612 (2003)
23.78 T. Niederkrotenthaler, T.E. Dorner, M. Maier: Development of a practical tool to measure the impact of publications on the society based on focus group discussions with scientists, BMC Public Health **11**, 588 (2011)
23.79 S. de Rijcke, P.F. Wouters, A.D. Rushforth, T.P. Franssen, B. Hammarfelt: Evaluation practices and effects of indicator use—A literature review, Res. Eval. **25**(2), 161–169 (2016)
23.80 Anon: The maze of impact metrics, Nature **502**(7471), 271 (2013)
23.81 M. Thelwall, K. Kousha: Web indicators for research evaluation. Part 2: Social media metrics, Prof. Inform. **24**(5), 607–620 (2015)
23.82 M. Thelwall, N. Maflahi: Guideline references and academic citations as evidence of the clinical value of health research, J. Assoc. Inf. Sci. Technol. **67**(4), 960–966 (2016)
23.83 R. Costas, Z. Zahedi, P. Wouters: Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective, J. Assoc. Inf. Sci. Technol. **66**(10), 2003–2019 (2015)
23.84 L. Bornmann, R. Haunschild, W. Marx: Policy documents as sources for measuring societal impact: How is climate change research perceived in policy documents?, Scientometrics **109**(3), 1477–1495 (2016)
23.85 R. Haunschild, L. Bornmann: How many scientific papers are mentioned in policy-related documents? An empirical investigation using web of science and altmetric data, Scientometrics **110**(3), 1209–1216 (2017)
23.86 N. Robinson-Garcia, D. Torres-Salinas, Z. Zahedi, R. Costas: New data, new possibilities: Exploring the insides of altmetric.com, Prof. Inform. **23**(4), 359–366 (2014)
23.87 L. Colledge: *Snowball Metrics Recipe Book* (Snowball Metrics Program Partners, Amsterdam 2014)
23.88 L. Bornmann, R. Haunschild: Which people use which scientific papers? An evaluation of data from F1000 and Mendeley, J. Informetr. **9**(3), 477–487 (2015)
23.89 L. Bornmann: Usefulness of altmetrics for measuring the broader impact of research: A case study using data from PLOS and F1000Prime, Aslib J. Inf. Manag. **67**(3), 305–319 (2015)
23.90 L. Bornmann: Alternative metrics in scientometrics: A meta-analysis of research into three altmetrics, Scientometrics **103**(3), 1123–1144 (2015)
23.91 M. Erdt, A. Nagarajan, S.C. Sin, Y.L. Theng: Altmetrics: An analysis of the state-of-the-art in measuring research impact on social media, Scientometrics **109**(2), 1117–1166 (2016)
23.92 L. Bornmann, R. Haunschild: How to normalize Twitter counts? A first attempt based on journals in the twitter index, Scientometrics **107**(3), 1405–1422 (2016)
23.93 L. Bornmann, R. Haunschild: Normalization of mendeley reader impact on the reader- and paper-side: A comparison of the mean discipline normalized reader score (MDNRS) with the mean normalized reader score (MNRS) and bare reader counts, J. Informetr. **10**(3), 776–788 (2016)
23.94 L. Bornmann, R. Haunschild: Measuring field-normalized impact of papers on specific societal groups: An altmetrics study based on Mendeley data, Res. Eval. **26**(3), 230–241 (2017)
23.95 R. Fairclough, M. Thelwall: National research impact indicators from Mendeley readers, J. Informetr. **9**(4), 845–859 (2015)
23.96 R. Haunschild, L. Bornmann: Normalization of Mendeley reader counts for impact assessment, J. Informetr. **10**(1), 62–73 (2016)
23.97 R. Haunschild, H. Schier, L. Bornmann: Proposal of a minimum constraint for indicators based on means or averages, J. Informetr. **10**(2), 485–486 (2016)
23.98 M. Thelwall: Three practical field normalised alternative indicator formulae for research evaluation, J. Informetr. **11**(1), 128–151 (2017)

23.99 R. Haunschild, L. Bornmann: Normalization of zero-inflated data: An empirical analysis of a new indicator family. In: *16th Int. Soc. Scientometr. Informetr. Conf. ISSI 2016, Wuhan* (2017), https://arxiv.org/abs/1704.02211

23.100 L. Bornmann, R. Haunschild: Normalization of Zero-Inflated Data: An Empirical Analysis of a New Indicator Family and its Use with Altmetrics Data, J. Informetr. **12**(3), 998–1011 (2018)

23.101 A. Patel, J. Chavda: Measuring research impact: Bibliometrics, social media, altmetrics, and the BJGP, Br. J. Gen. Pract. **66**(642), E59–E61 (2016)

23.102 L. Bornmann, R. Haunschild: To what extent does the Leiden manifesto also apply to altmetrics? A discussion of the manifesto against the background of research into altmetrics, Online Inf. Rev. **40**(4), 529–543 (2016)

23.103 S. Haustein, T.D. Bowman, K. Holmberg, A. Tsou, C.R. Sugimoto, V. Lariviere: Tweets as impact indicators: Examining the implications of automated "bot" accounts on twitter, J. Assoc. Inf. Sci. Technol. **67**(1), 232–238 (2016)

23.104 J.A. Arevalo, J.A. Cordon-Garcia, B.M. Barba: Altmetrics: Measuring the influence of the media on the social impact of research, Cuad. Doc. Multimed. **27**(1), 75–101 (2016)

23.105 S. Haustein: Grand challenges in altmetrics: Heterogeneity, data quality and dependencies, Scientometrics **108**(1), 413–423 (2016)

23.106 J. Callaert, B. Van Looy, A. Verbeek, K. Debackere, B. Thijs: Traces of prior art: An analysis of non-patent references found in patent documents, Scientometrics **69**(1), 3–20 (2006)

23.107 K. Kousha, M. Thelwall: Patent Citation Analysis with Google, J. Assoc. Inf. Sci. Technol. **68**(1), 48–61 (2017)

23.108 F. Narin, K.S. Hamilton, D. Olivastro: The increasing linkage between US technology and public science, Res. Policy **26**(3), 317–330 (1997)

23.109 C.K.N. Chang, A. Breitzman: Using patents prospectively to identify emerging, high-impact technological clusters, Res. Eval. **18**(5), 357–364 (2009)

23.110 F. Narin, M.P. Carpenter, P. Woolf: Technological performance assessments based on patents and patent citations, IEEE Trans. Eng. Manag. **31**(4), 172–183 (1984)

23.111 Elsevier Research Intelligence: SciVal Patent Metrics, https://www.elsevier.com/research-intelligence/resource-library/scival-patent-metrics (2016)

23.112 J.M. Grimshaw, I.T. Russell: Effect of clinical guidelines on medical-practice—A systematic review of rigorous evaluations, Lancet **342**(8883), 1317–1322 (1993)

23.113 P. Shekelle, S. Woolf, J.M. Grimshaw, H.J. Schunemann, M.P. Eccles: Developing clinical practice guidelines: reviewing, reporting, and publishing guidelines; updating guidelines; and the emerging issues of enhancing guideline implementability and accounting for comorbid conditions in guideline development, Implement. Sci. (2012), https://doi.org/10.1186/1748-5908-7-62

23.114 J.P. Andersen: Association between quality of clinical practice guidelines and citations given to their references. In: *14th Int. Soc. Scientometr. Informetr. Conf., ISSI 2013, Vienna* (2013)

23.115 M. Thelwall, N. Maflahi: Guideline references and academic citations as evidence of the clinical value of health research, J. Assoc. Inf. Sci. Technol. **67**(4), 960–966 (2015), https://doi.org/10.1002/asi.23432

23.116 D. Kryl, L. Allen, K. Dolby, B. Sherbon, I. Viney: Tracking the impact of research on policy and practice: Investigating the feasibility of using citations in clinical guidelines for research evaluation, BMJ Open **2**(2), 7 (2012)

23.117 G. Lewison, R. Sullivan: The impact of cancer research: How publications influence UK cancer clinical guidelines, Br. J. Cancer **98**(12), 1944–1950 (2008)

23.118 T.H. Jones, S. Hanney: Tracing the indirect societal impacts of biomedical research: Development and piloting of a technique based on citations, Scientometrics **107**(3), 975–1003 (2016)

23.119 F. Bunn, D. Trivedi, P. Alderson, L. Hamilton, A. Martin, E. Pinkney, S. Iliffe: The impact of Cochrane reviews: A mixed-methods evaluation of outputs from Cochrane review groups supported by the national institute for health research, Health Technol. Assess. (2015), https://doi.org/10.3310/hta19280

23.120 A. Ritter, K. Lancaster: Measuring research influence on drug policy: A case example of two epidemiological monitoring systems, Int. J. Drug Policy **24**(1), 30–37 (2013)

23.121 J. Liu: New Source Alert: Policy Documents, http://www.altmetric.com/blog/new-source-alert-policy-documents/ (2014)

23.122 T. Arthur: Categorizing Policy Document Citations in PlumX, http://plumanalytics.com/categorizing-policy-document-citations-in-plumx/ (2016)

23.123 A. Ballabeni, A. Boggio, D. Hemenway: Policies to increase the social value of science and the scientist satisfaction. An exploratory survey among Harvard bioscientists, F1000Research **3**, 20 (2014)

23.124 L. Bornmann: Is there currently a scientific revolution in scientometrics?, J. Assoc. Inf. Sci. Technol. **65**(3), 647–648 (2014)

23.125 L. Bornmann: Scientific revolution in scientometrics: The broadening of impact from citation to societal. In: *Theories of Informetrics and Scholarly Communication*, ed. by C.R. Sugimoto (De Gruyter, Berlin 2016) pp. 347–359

Lutz Bornmann

Division for Science and Innovation Studies
Administrative Headquarters of the Max Planck Society
Munich, Germany
bornmann@gv.mpg.de

Lutz Bornmann works as a Sociologist of Science at the Division for Science and Innovation Studies of the Max Planck Society in Munich. His current research interests include research evaluation, peer review, bibliometrics, and altmetrics. He is a member of the editorial board of the *Journal of Informetrics* (Elsevier), *PLOS ONE*, and *Scientometrics* (Springer). Furthermore, he is a Senior Associate Editor at theInternational Journal of Biomedical Science Editing (Kowsar).

Robin Haunschild

Central Information Service
Max Planck Institute for Solid State Research
Stuttgart, Germany
r.haunschild@fkf.mpg.de

Robin Haunschild studied Chemistry in Hannover and Marburg. After two Postdoctoral positions (Rice University in Houston and Karlsruhe Institute of Technology) he joined the Central Information Service for the Institutes of the Chemical, Physical, and Technical Section of the Max Planck Society in 2014. His current research interests include the study of altmetrics and the application of bibliometrics to specific fields of natural sciences, e.g., climate change.

24. Econometric Approaches to the Measurement of Research Productivity

Cinzia Daraio

The measurement of research productivity is receiving more and more attention. Besides scholars that are interested in understanding how research works and evolves over time, there are supranational, national and local governments, and national evaluation agencies, as well as various stakeholders, including managers of academic and research institutions, scholars and more generally the wider public, who are interested in the accountability and transparency of the scholarly production process.

The main objective of this chapter is to analyze econometric approaches to research productivity and efficiency, highlighting what econometric approaches to research assessment can offer and what their benefit is, compared to traditional bibliometric or informetric approaches. We describe the nature of, and the ambiguities connected to, the measurement of research productivity, as well as the potential of econometric approaches for research measurement and assessment. Finally, we propose a checklist when developing econometric models of research assessment as a starting point for further research.

The main objective of this chapter is to describe what econometrics can offer to informetrics. It is a set of conceptual and methodological models with estimation tools to provide empirical estimates of research productivity and efficiency (towards effectiveness and impact) conceived as a multidimensional and complex related set of activities. The chapter relies on [24.1], which is its departure base, and on [24.2–4].

As we will see in the next sections, the evolution of research in economics and econometrics related to the measurement of productivity and efficiency in addition to the changes in research activity itself is challenging the traditional efficiency estimation methods.

We think that due to the complex nature of the research activity, *nonparametric methods* (which do not rely on strict hypotheses about the parameters of the re-

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_24

lationships among inputs and outputs) can offer a valid analysis tool, with a broader generality than *parametric methods* (which in turn are based on more restrictive hypotheses about the parameters). Nevertheless, the *validity* and *appropriateness* (quality in our framework) must be conceptualized, formalized and analyzed with reference to the specific problem (fitness for purpose). Within this context, *semiparametric* and parametric methods can also be useful, perhaps in combination with nonparametric methods used as preliminary exploratory methods.

We do not have a specific approach or technique that dominates all the others in every context of application, and there are a lot of methods that are proposed in the specialized literature each day. We have hence to choose the estimation tool according to the problem at hand. There are many tools and there are many problems and there are many different fields of research that study the problems through different perspectives. The selection of the most appropriate tool is not easy given also the rapid evolution, the technicalities and jargon of the different disciplines. This rapid evolution currently affects both the research activity and the productivity estimation. In order to be able to understand which may be the most suitable tool, to highlight pros and cons and the consequent caveats for the users of the estimation we propose a very rough checklist.

24.1 Assessing the Productivity of Research

The quantitative assessment of the productivity of the research activity is not only a fascinating and complex research problem, but it is also a relevant policy issue. This is because it may be connected to policies of evaluation, funding allocation, promotion and more generally to performance comparisons, which can be carried out at different level of analysis, such as at a macro (country), meso (regional), micro (institutional) and micro-micro (individual and group-based) level.

In fact, the measurement of productivity of research is receiving more and more attention by different policymakers. Beside scholars that are interested in understanding how research works and evolves over time, there are supranational, national and local governments, and national evaluation agencies, as well as various stakeholders, including managers of academic and research institutions, scholars and more generally the wider public, who are interested into the accountability and transparency of the scholarly production process.

The main objective of this chapter is to analyze econometric approaches to research productivity and efficiency, highlighting what econometric approaches to research assessment can offer and what their benefit is, compared to traditional bibliometric or informetric approaches. We will analyze the nature and potential of econometric approaches for research measurement and assessment. In particular we will show what the potential benefits of nonparametric techniques and more advanced econometric frontier models are.

Before entering into its purposes, the chapter describes the main concepts involved and illustrates the main ambiguities connected to the measurement of research productivity. In particular, we will show that the measurement of research productivity is a *component* of research assessment and its ambiguities are essential and should be taken into account. We will then introduce the need for a framework to accomplish this task and will show some examples of its usefulness in research assessment.

In recent decades, the rapid changes taking place in the production, communication and evaluation of research have been signs of an ongoing transformation. One could maintain that we are living a sort of middle age guided by the information and communication technologies (ICT) revolution, or the so-called *fourth revolution* as described by *Floridi* [24.5], which emphasizes the importance of information.

Largely, the current middle age of research evaluation might be understood as the transition from a traditional evaluation model, based on bibliometric indicators of publications and citations, to a modern evaluation, characterized by a multiplicity of distinct, complementary dimensions. This step is guided by the development and increasing availability of data and statistical and computerized techniques for their treatment, including among others the recent advancements in artificial intelligence and machine learning.

At the same time, the recent changes also have an impact on econometric techniques and therefore on the *ways* in which production is conceived (and represented) and productivity measured (estimated). It will therefore be necessary to present all these changes in a synthetic way before analyzing the potential of the econometric techniques for the evaluation of research. The main methodological challenges related to estimating productivity and efficiency will be made clear to the reader and analyzed in detail.

Finally, this chapter applies the notions of standardization and harmonization to the methodological dimension of the research assessment. We are well aware that this is a difficult and challenging task because, among other factors, it depends on the purposes of the assessment. Nevertheless, the proposed checklist when developing econometric models of research assessment, although very preliminary, might constitute a useful starting point for further research.

24.2 What Do We Measure?

24.2.1 Some Hints from a Recent Debate

The theme of the evaluation of research productivity may be subject to easy misunderstandings, as has been clearly shown by *Glänzel* et al. [24.6] in the debate on [24.7]. "Mean normalized citation scores (MNCSs) and like size-independent indicators". In this debate, different terms were used, such as *research performance*, *productivity*, *efficiency*, and *true evaluation of research productivity* [24.8] without an in-depth introduction and discussion of their meanings.

We do not enter here into the discussion about the MNCS, which relates to *output bibliometrics* [24.9]. Further information about MNCS can be found in Chap. 11 in this Handbook.

Nevertheless, we take into account some of the warnings provided within this debate for the development of this chapter. In particular, we refer to the contributions of *Zitt* [24.9] and *Glänzel* et al. [24.6]. *Zitt* has observed that [24.9, p. 675]:

> More generally, economics of science (as sociology of science) rightly claims science networks, the core of scientometrics, as important objects, with two implications: 1) Output bibliometricians should remain modest and prudent, and welcome strong concepts and the toolbox coming from economics. On the policy side, many bibliometricians have devoted much time to warn against the limitations of bibliometric measures, to emphasize their properties and their conditions of use, to stress the dangers of misusage, and to insist on the traps of one-figure descriptions. 2) Conversely, authors with another background should be cautious in the criticism of bibliometric measures considered independently from their historical context or the question of sources and data.

Glänzel et al. [24.6, pp. 658/659], on the other hand, have pointed out a series of important aspects to consider. They include the question of the *correct interpretation*, the fact that *productivity and performance* are distinct concepts; the *multidimensionality* of performance; and the distinction between the *efficiency* and the *effectiveness* and the observation that system performance relates to the effectiveness of a system, while productivity only hints at its efficiency. *Efficiency or productivity* also poses its specific challenges in terms of operationalization and quality of the underlying data; *comparability (and hence validity) of the underlying data themselves* constitutes *a problem*, and finally the authors addressed the *difficulty of quantification*.

24.2.2 The Measurement of the Productivity as a Component of a Research Assessment

Before turning to the definition of the basic concepts, a distinction is in order here. There are at least two main reasons for carrying out a measurement of the productivity of research. The first one (study purpose) is to understand how research is produced, how research works, and its evolution over time. The second (assessment purpose) is related to the use of the measurement of productivity carried out as an assessment measure in an evaluative framework of the research activity. We should always take into account why we are carrying out our measurement, because given the peculiarity of the research activity, in the case of an assessment purpose much more care should be taken. Generally, we cannot speak about *true measurement of research efficiency* in either cases, because, as will be shown in the continuation of this chapter, the assessment of productivity and efficiency is *tricky* and involves several conceptual, analytical, methodological and data-related *issues*. Being more specific, it is much more dangerous to speak about *true measurement of research efficiency* if the assessment is done in an evaluative framework in which some scholars are compared with each other for the allocation of research funds, than in a research paper that analyses the scientific developments of a set of countries. In our opinion, the analyst should be careful in both cases, and also in the second case should state clearly the underlying assumptions, and the simplistic notions that were made to handle a much more complex issue. Nevertheless, as far as the first case is concerned, it is important to warn the users of the productivity/efficiency measurement about all the hypotheses, simplifications and possible distortions, because in this case, the measurement may affect individual people and society at large much more than a wrong or a superficial paper. This is linked to the distinction between *research tools* and *management or assessment tools*, well known in the quantitative studies of science [24.10, p. 90].

Measuring is different from assessing, which in turn is different from evaluating. According to the Cambridge dictionary (https://dictionary.cambridge.org/), to measure means *to discover the exact size or amount of something*, to assess means *to judge or decide the amount, value, quality, or importance of something*, and to evaluate means *to judge or calculate the quality, importance, amount, or value of something*. Assess and evaluate are then considered as synonyms. Nevertheless, for the sake of clarity, we adopt the distinction made by *Moed* [24.10, p. 94] and will consider, in the following, assessment as "the total of activities in assessment or evaluation processes, or the act of evaluating or assessing in general". The research assessment, according to *Moed* [24.10, Table 6.3, p. 95], includes four domains, namely:

1. Policy or management
2. Evaluation
3. Analytics
4. Data collection.

The *policy or management* formulates the policy issue and the assessment objectives, it defines the organizational aspects and the budget and its outcome is to make a policy decision based on the outcomes from the evaluative domain. The *evaluation* defines and evaluates worth, it specifies the *evaluative framework* that is a set of evaluation criteria in agreement with the constituent policy issue and assessment objectives, and its outcome is making judgments on the basis of the evaluative framework and the empirical evidence collected.

This distinction is important here because it allows us to clearly understand that the measurement of the productivity of research is a *component* of a research assessment. For this reason, when research productivity measurements are carried out within a research assessment, we have to take into account the assessment context in its *entirety*.

24.2.3 What is Research?

To define the research, we propose to the readers the definitions of the Frascati manual developed by the Organisation for Economic Co-operation and Development (OECD), which are *definitions for measurement purposes* and represent a commonly shared standard in the literature.

In the *Frascati manual* last edition [24.11] the basic definitions of research and development (R&D) are slightly modified with respect to the previous edition of the manual. Some extensions of the current version of the manual refer to the description of the peculiarities of R&D in social science and humanities [24.11, pp. 55–57]. In addition, the 2015 edition of the Frascati manual introduces a set of five core criteria that explain what key features must be met for an activity to qualify as R&D. Table 24.1 reports the definition of R&D and its subcomponents together with the five criteria adopted by the last edition of the Frascati manual.

The definition of R&D subcomponents reported in Table 24.1 is consistent with the definition of R&D used in the previous 2002 edition of the Frascati manual and covers the same range of activities.

From the definitions reported in Table 24.1, it follows that the research productivity measured by means of any analytic method and/or empirical approach, even the most sophisticated one, should be considered as an *estimate* of the *true* research productivity, which is unknown and very difficult to assess.

24.2.4 What is Productivity?

Basic Notions

The topic of productivity has been at the core of the economic analysis since Adam Smith's *Wealth of Nations* pin factory, and even before. *Productivity* is commonly defined as a ratio between the output produced and the inputs used to produce it, in a given production process. The measurement of research productivity can then benefit from the application of methods developed in this field. The measurement of the productivity of an economic activity can be carried out in different ways and according to different approaches.

A handbook chapter by *Bonaccorsi* and *Daraio* [24.1], which is the starting base of this chapter, carried out a survey and discussed the potential and limitations of econometric methods for the evaluation of the productivity of scientific and technological systems. They compared the advantages and disadvantages of the main approaches proposed in the literature, namely the production function and the production frontier (or efficiency analysis) approach.

We provide here a short overview of the main methods, starting with a brief sketch of ratio measures and index numbers. This is taken from [24.1]. After that, we will analyze the two main approaches proposed in the literature, namely production functions and production frontiers (or efficiency analysis).

A very simple *measure of productivity* is given by the ratio between the output realized and the input used to produce it. This measure considers one category of input and relates it to one category of output, without considering the complementarity and substitution rela-

Table 24.1 Definition of R&D and its subcomponents according to the *Frascati manual* last edition [24.11]

Concept	Definition	Page in the Frascati manual (2015)
Important premise about R&D	The defining feature of R&D in this manual is that it is carried out in order to generate new knowledge as an output, irrespective of its purpose, which could be the generation of economic benefit, addressing societal challenges or simply having the knowledge in itself. This intentionality is used in this manual to distinguish between experimental development and basic and applied research. In the same manner, it is of interest to identify and, if possible, to measure the different types of outputs from R&D. However, it is difficult to identify and measure R&D outputs. This is due to a series of factors that affect both how knowledge is distributed and used in the economy and the complementary inputs necessary for results to occur. Any outputs and effects may take a long time to be realized and may occur at different places and for different actors than those carrying out the R&D. *Only very partial outputs can currently be directly identified and measured* as part of collecting information on R&D activities and funding.	*OECD* [24.11, p. 25]
R&D definition	Research and experimental development (R&D) comprises creative and systematic work undertaken in order to increase the stock of knowledge—including knowledge of humankind, culture and society—and to devise new applications of available knowledge.	*OECD* [24.11, p. 44]
Core criteria to identify R&D	For an activity to be an R&D activity, it must satisfy five core criteria. The activity must be: 1. Novel 2. Creative 3. Uncertain 4. Systematic 5. Transferable and/or reproducible.	*OECD* [24.11, p. 45, see also pp. 46–48]
Definition of subcomponents of R&D (basic research, applied research and experimental development)	The term R&D covers three types of activity: basic research, applied research and experimental development. *Basic research* is experimental or theoretical work undertaken primarily to acquire new knowledge of the underlying foundation of phenomena and observable facts, without any particular application or use in view. *Applied research* is original investigation undertaken in order to acquire new knowledge. It is, however, directed primarily towards a specific, practical aim or objective. *Experimental development* is systematic work, drawing on knowledge gained from research and practical experience and producing additional knowledge, which is directed to producing new products or processes or to improving existing products or processes.	*OECD* [24.11, p. 45], [24.12]
Product	The manual follows the system of national accounts (SNA) convention in which *product* refers to a good or a service.	*OECD* [24.11, p. 45]
Process	Refers to the transformation of inputs to outputs and to their delivery or to organizational structures or practices.	*OECD* [24.11, p. 45]

tions between inputs, and ignoring the effects of joint production in outputs. They may be used as a kind of first-order approximation.

These ratio-based measures of productivity (output–input) are defined also as *partial* productivity measures (see *Daraio* and *Simar* [24.13, p. 14], for their description within the efficiency analysis context). On the other hand, *total* factor productivity measures aim at the measurement of a value of the output–input ratio, which considers *all* outputs and inputs. The estimation of total factor productivity measures by combining all inputs and all outputs to obtain a single ratio aims to avoid imputing gains to one factor (or one output) that should be attributed to some other input (or output). However, total factor productivity measures present aggregation problems such as choosing the weights to be used in order to obtain a *single output to single input* ratio.

Another simple measure of productivity is given by *index numbers*. An index number is defined as a real number that measures changes in a set of variables. In particular, index numbers are applied to measure price and quantity changes over time, as well as to measure differences in the levels across firms, industries, regions, or countries. Panel data allow the measurement of productivity change as well as the estimation of technical progress or regress. Productivity change occurs when an index of outputs changes at a different rate from that at which an index of inputs does. Productivity change can be calculated using index number techniques such as Fischer or Tornqvist productivity indices. Both these indices require quantity and price information, as well as assumptions about the structure of the technology and the behavior of producers. Productivity change can also be calculated using a production frontier approach to construct a Malmquist productivity index. This approach does not require price information or technological and behavioral assumptions, and allows the identification of the sources of measured productivity change (i. e., technological progress/regress and efficiency changes). It requires the estimation of a representation of production technology that can be made using different frontier approaches (see *Daraio* and *Simar* [24.13, p. 14], for their description within the efficiency analysis context).

A distinction between productivity and efficiency is in order here. Although in the specialized productivity and efficiency analysis literature the terms productivity and efficiency are used often as synonyms, we think it is more appropriate to distinguish their meanings [24.13].

Productivity is the ratio of the outputs over the inputs, and we may consider it as an efficiency *in the narrow sense*. Efficiency, in the broad sense, is defined instead as the output–input with respect to an estimated reference frontier, or frontier of the best practices [24.13, p. 14]. Given that we have introduced the concept of frontier, we can analyze the differences between production functions and production frontiers in the next section.

Production Functions versus Production Frontiers

As recalled above, the two main approaches proposed in the literature to econometrically estimate productivity and efficiency are production functions and production frontiers (or efficiency analysis). In the production function approach, the measurement of productivity relies on an average relation or a relationship that intersects the input–output data, looking for the *representative* units (or expected behaviors). Most often, the estimation of production function is based on the specification of a functional form of the relationship between inputs and output, characterized by coefficients that relate the inputs to the output (which is typically univariate).

In any discipline, knowing the past is useful to understand the present and the current challenges. The knowledge and history of production functions may therefore be informative for the main purpose of this chapter. *Mishra* [24.14] has carried out a short description of the evolution of the concept and econometrics of production functions. He highlights the assumption of technical efficiency behind the production function, its main economic features and the subsequent controversies in economics about the use of aggregate production functions.

Thanks to the parameters of the production function, from which the name parametric comes from, it is possible to analyze the level of productivity, which is usually given by a coefficient that multiplies the function (this is the case of neutral technical progress); the marginal productivity of each factor (making the assumptions that the factors can be measured without ambiguity, the other inputs can be kept constant, and the availability of an infinite number of techniques such that the passage from one combination of factors to another could happen also for infinitesimal variations); the marginal rate of substitutions amongst factors; the factors' intensity, given by the ratio of the amount of two inputs, given the marginal rate of substitutions; the optimal choice of the combination of inputs, through the equality of the factors' marginal rate of substitutions and their prices ratio; simple measures of productivity by obtaining the ratio of the observed level of output over the production function optimal level; measures of technical change; returns to scale; inputs' elasticity of substitution, and the like.

One may also estimate more generally, without relying on parametric assumptions, i. e., nonparametrically, a production function, by applying methods from nonparametric statistics and econometrics (see basic definitions in Table 24.4, and *Hardle* [24.15], and *Henderson* and *Parmeter* [24.16]. This has not been done very often in the empirical literature. One reason is the economic interpretation of coefficients that facilitate the understanding of the data, and another reason is that up to recently there was not a lot of econometric software implementing nonparametric techniques.

In the *production frontier* approach, or efficiency analysis, the measurement of productivity is based on the estimation of a frontier that envelops the data, and in gaging the distance of *each unit* from the estimated *best performing* frontier.

The production frontier approach or efficiency analysis ([24.17, 18] and [24.13, p. 1]) challenges the basic implicit assumption of production function efficiency. Indeed, in standard microeconomic theory [24.19], a production function describes the maximum achievable output for given inputs. With respect to input–output characterization, production functions then implicitly assume a *maximizing* behavior of the units.

A rigorous analytical approach to productivity and efficiency originated with the works of *Koopmans* [24.20], *Debreu* [24.21] and *Shephard* [24.22], among others, and was empirically applied by *Farrell* [24.23], *Charnes* et al. [24.24] and others. These seminal works gave rise to a considerable amount of studies that challenged theoretical mainstream production analysis focused on production activity as an optimization process.

One may specify and estimate a production frontier by means of a *parametric* production frontier (in which the relationship between input–output relies on specific parameters of their functional form) or through a *nonparametric* production frontier, based for instance on linear programming techniques that envelop the data without assuming any functional relations between input–output. Data envelopment analysis (DEA) [24.23, 24] is a nonparametric approach based on linear programming techniques.

Mishra presents the nonparametric approach to efficiency analysis, based on DEA as a more realistic approach, as follows [24.14, p. 7]:

> It has been noted that although the notion of production function generally assumes that technical efficiency has been achieved, this is not true in reality. Some economists and operations research scholars [...] addressed this problem by what is known as the Data Envelopment Analysis (DEA). The advantages of DEA are: here one need not specify a mathematical form for the production function explicitly; it is capable of handling multiple inputs and outputs and being used with any input/output measurement; and efficiency at technical/managerial level is not presumed. It has been found useful for investigating the hidden relationships and causes of inefficiency.

On the other hand, most empirical production analysis has used the production functions and has focused on a central tendency, or *average* or *most likely* relationship constructed by intersecting data with a parametric function. Estimating production functions has been carried out by estimating the coefficients of regression equations, which describe the average tendency of the relationship between inputs and outputs. But as we have seen above, nonparametric specifications of production functions are also available. The purpose of efficiency analysis, based on frontiers, is to make a relative benchmark or comparison among units. Production frontiers, as well, may be specified by means of a *parametric* functional form for the frontier or *nonparametrically*, i. e., without a functional specification of the frontier.

Another useful distinction of frontier models is between *deterministic* and *stochastic* frontier models. In deterministic frontier models, such as DEA, all the deviation from the estimated production frontier is assumed to be inefficiency. There is no noise in this kind of model. On the other hand, stochastic frontier models, or stochastic frontier analysis (SFA) [24.25], assume there is some noise in the model. SFA models distinguish noise from inefficiency, but to do this they rely on the parametric specification of the functional form for the noise and on the functional form for the inefficiency. These functional specifications are added to the functional form specified for the production frontier. There are in the literature also *semiparametric* approaches that propose models in which some functional forms are parametrically specified and some others are left nonparametric. From the empirical point of view, the efficiency analysis approach offers techniques for estimating the *efficient* production frontier and for measuring and interpreting the relative efficiency of each unit as a distance with respect to this estimated frontier. We will come back to these distinctions in Sects. 24.6 and 24.7 highlighting their methodological challenges and their potential.

Some Features of Research Productivity from the Economics of Science

Let us focus mainly on scientific production in the public sector research system. The first part of this section is taken from [24.1]. Scientific production is not only a multi-input multi-output process, but the relation between inputs and outputs is nondeterministic, uncertain, lagged, nonlinear, and subject to important but subtle external effects.

We know from the *economics of science* [24.26] that a few stylized facts about individual productivity do exist [24.27]. First, the distribution of individual productivity of scientists is extremely skewed, with a small percentage of very productive scientists accounting for a disproportionate share of publications. Second, productivity declines over a scientist's life cycle [24.28]. Gender effects on scientific productivity [24.29, 30] have been recently investigated [24.31, 32] together with collaboration [24.33, 34] and funding [24.35]. These very basic features of scientific production make a representation in which the marginal rate of substitution between units of inputs is constant or independent on size, and in which interaction effects are zero, which is highly unrealistic.

How these individual-level factors combine on an organizational and institutional level is, in fact, a very open question [24.36]. Do people with the same individual productivity attract each other, or perhaps are hired according to a consistent quality strategy, so that in the end the same skewed distribution will also be observed across organizations and institutions? Or, quite to the contrary, do people with different individual productivities mix within research departments and institutes? What is the effect of the organizational setting on individual productivity? An important effect of the external environment of scientists has also been observed, in terms of complementary resources, time constraints, and social incentives at the level of department or institute [24.37–41]. External factors may create complementarities, which have a nonlinear effect. Under these conditions, the lack of a specification of the functional form is a clear advantage.

Research activities are intrinsically multi-output activities.

First of all, for a large part of the research system the allocation of the time of researchers takes place between research and teaching. Since the share of time is not fixed across disciplines and countries, it is sensible to take both outputs into consideration, when possible.

Second, within the narrow area of research, whilst the single most important output is clearly scientific publications, it is difficult to claim that other outputs such as patents, software, advisory work for the government, consulting, or technical assistance do not have any relevance with respect to research.

Finally, scientific publications cover a large range of specific outputs, such as papers in refereed journals, papers in technical or professional journals, notes, reviews, books, and edited books. How much worth is a book with respect to a paper in a refereed journal? Do more papers in the technical press compensate for fewer papers in academic journals?

In order to take into consideration the multi-output nature of research it is necessary to aggregate each type of output. Using a multi-output specification is clearly more appropriate.

The economics of science [24.42] reminds us that researchers conduct research for different reasons including their interest in *puzzle solving*, reputation based on the priority of their discovery, awards and recognition for their achievements, and also through publications, which has a key role for funding and promotion. Other incentives vary across fields. Research is a public good. This means that once it is public others cannot be excluded from its use (nonexcludable nature). This offers the possibility of *free-riders* behaviors and difficulty in capturing economic returns. Moreover, once public goods are produced, competitive markets fail to provide them efficiently. In research, reward is not for effort but for achievement.

What the economics of science tells us about the production of scientific research is that research does not only produce multiple *outputs*, but also involves multiple *inputs*, including knowledge, time, materials and equipment. Some inputs are embedded in people (knowledge and time in particular) and most of these inputs are expensive.

As observed by *Stephan* [24.42, p. 228]:

> Incentives and cost matter for science and economics is also about the allocation of scarce resources across competing wants and needs, economics is also about whether resources are allocated efficiently.

When concluding her book on *How Economics Shapes Sciences*, *Stephan* [24.42, p. 235] states three more general and difficult *efficiency* (in the narrow sense) questions that require further research. These are:

1. What is the right amount of money in terms of percent of the gross domestic product (GDP) to spend on university R&D?
2. What is the most efficient mix in terms of budget allocated to the different disciplines?
3. What is the most efficient structure for grants in terms of size, duration, criteria for evaluation and number of people involved?

Productivity is Just a Component of Performance

Generally, in this chapter, by performance we mean the results obtained by an activity. *Bazeley* [24.43], developing the work of *Åkerlind* [24.44], provides an attempt to conceptualize what research performance is. This is useful for our chapter showing that research productivity is just a component of research performance, which relates mostly to the output produced, but it does not coincide with research performance, which is a much broader concept.

In conceptualizing research performance *Bazeley* [24.43] identifies two basic components (research activity and making research visible), with six secondary-level dimensions. The four essential dimensions of the research activity (or components of research performance), identified by *Bazeley* [24.43] were:

1. *Engagement* (interest and involvement)
2. *Task orientation* (disciplined management, getting the job done)
3. *Research practice* (knowledge and skill, substantively and methodologically sound)
4. *Intellectual processes* (analytic capacity and creative thinking).

Two other dimensions (of which at least one is necessary) relating to the performance, or making research visible, component of research performance were *dissemination* (formal communication of research outcomes) and *collegial engagement* (sharing knowledge and expertise). Research performance was in addition seen to occur within conditions provided by an institutional context (education and training; opportunity and resources), and to bring about a range of outcomes (product, impact and reputation). Table 1 of [24.43, pp. 895/896] lists eight characters of researcher attributes: quality, ability, *productivity*, recognition, benefit, activity, satisfaction and approachability. It appears again, clearly, that productivity is just one component of research performance.

A simplistic economic view that considers research activity as any other economic activity and measures it through a basic production function, calculates simple partial productivity measures [24.8, 45]; although we do not share this approach, we admit that in certain contexts and for certain evaluation questions, with all the caveats of the case, it could be useful. What is important, in this case, is to clearly warn the users of the productivity measurement about the underlying assumptions, the strong simplifications and about the bias existing in using this measure in a research assessment. Much more caution should be added when individuals are the entity subject to the assessment, for which we know all the problems related to individual-level bibliometrics (see an outline in [24.46]) and ambiguities that bibliometric indicators may mask in assessing individual productivity [24.47]. An indicator, according to the Cambridge dictionary, is something that shows what a situation is like. *Van Raan* defines an indicator as [24.48, pp. 21/22]:

> The result of a specific mathematical operation (often simple arithmetic) with data. The mere number of citations of one publication in a certain time period is data. The measure in which such citation counts of all publications of a research group in a particular field are normalised to citation counts of all publications worldwide in the same field, is an indicator. An indicator is a measure that explicitly addresses some assumption.

24.2.5 Ambiguities of the Measurement of Research Productivity

Bonaccorsi and *Daraio* [24.1] discussed a general concept of *research productivity*. It is based on the relationship which exists between a set of inputs (or resources) used to realize a set of outputs (or products) in a given production process. In contrast to a standard production activity, the production process of the research activity is seen as being characterized by several conceptual and measurement problems, which affect any definitional elements of the productivity, namely the inputs, the outputs and their functional relation. Indeed, a research activity may be represented as a process, characterized by a multiple inputs–multiple outputs relation, in which both the inputs and the outputs are heterogeneous and sometimes incommensurable, the inputs–outputs relation is nondeterministic, and the output is lagged, but with a nonfixed lag structure.

As we will see in the remainder of this chapter (in particular in Sect. 24.6), the current changes in research assessment (that are discussed in the next section) have amplified the ambiguities and the problems related to the measurement of research productivity.

24.3 Research Assessment in the Current Time and the Need for a Framework

To understand and tackle the challenges of the measurement of research productivity we need to analyze the changes taking place both in the realization of the research activity and in the econometric measurement of productivity. In this section, we analyze the former, while the next section describes the changes applying to the latter.

24.3.1 A Simplified Overview

Table 24.2 summarizes the main trends in research assessment exploring a categorization in factors in terms of change and consequences produced on the research assessment. This categorization is just a way to provide an outline, without claiming to be either complete or robust. In fact, some changes may be also consequences and some consequences may be changes.

As observed by *Nowotny* et al. [24.50]:

> The emergence of more open systems of knowledge production (Mode-2 Science) and the growth of complexity and uncertainty in society (Mode-2 Society) are phenomena linked in a co-evolutionary process.

The changes in knowledge production has led to the so-called *postacademic science* [24.49, 50]. The context speaks of changing science. The emergence of mode-2 society raises acute issues of social justice, economic equality and the further democratization of

Table 24.2 A simplified overview of the current situation in research assessment (based on [24.4, Table 1])

Category	Description
Changes	Has changed the way in which the knowledge is produced, the dynamics of science and its interactions with society: *postacademic science* [24.49, 50]
	There is a crisis of *technoscience* (scientific research and technological innovation, focused on applications [24.51]) and science identified [24.52] in reproducibility, peer-review, publication metrics, scientific leadership, scientific integrity and the use of science for policy
	Advent of the big data era and its technological developments in research assessment (the *computerization of evaluative informetrics*) [24.10]
	Has changed the way in which science is communicated [24.53]
Consequences or effects	*On the demand side* (those that ask for research assessment): changes of the requests and the ways in which the assessment is carried out (has to be done): 1. Extension to societal value and value for money (*evaluation society*) [24.54, 55] 2. Performance-based funding [24.56, 57] 3. Requests for new and timely indicators in response to changing needs [24.58] 4. Increase of institutional and internal assessments.
	On the supply side (those that offer research assessment): proliferation of rankings (among many others [24.59]), development of Altmetrics [24.60, 61], open-access repositories [24.62, 63], new assessment tools—both commercial (InCites and Sci-Val) and freely available (Google Scholar citation), desktop bibliometrics ([24.64]; Publish or Perish software)
	On scholars: the increase of *publish or perish* pressure, impact on the incentives, behavior and misconduct, and increasing critics against traditional bibliometric indicators [24.65–73]
	On the assessment process: Increasing *complexity of the research assessment* linked to the *implementation problem* [24.2]; multidimensionality of the assessment of the research [24.74]; problems of data quantification, harmonization and standardization for different evaluation and assessment purposes [24.75–77]
	On the measurement of productivity/efficiency within an assessment process: The increasing complexity of the research assessment and the extension of the boundaries of the research activity and the interdependence with the society requires a more precise description and delineation of the *boundaries* of the production process whose productivity has to be measured before making the estimate, and to consider the *dynamics* of the inputs, outputs and their connection

knowledge. Contextualization means that the implications as well as the applications of scientific research have to be embraced. Contexts emerge in relation to particular problems for which they are or may become relevant. Contextualization means that the human element and subjective experience is seriously taken into account. Researchers have to use social knowledge. The more open and comprehensive the scientific community, the more *socially robust* will be the knowledge it produces. Producing reliable knowledge is not sufficient. Socially robust knowledge is requested. The reinterpretation of the boundary of the production of knowledge is essential to describe and represent the production process. As a matter of fact, science enters the agora, and the range of perspectives in the agora increases: there is an increasing role of the construction of narratives of expertise to deal with this complexity and uncertainty. Knowledge is distributed, contextualized and heterogeneous. *Bucchi* analyzed the crisis of the technocratic view of science and the emergence of participatory processes and consider these as [24.51, p. ix]:

> Epochal changes in the social role of science, and generally in the production of scientific knowledge, and such changes concern the nature itself of contemporary politics and democracy.

The increasing intersection and permeability of the boundaries between science and society impacts on the dynamics of science communication and challenges it, as illustrated by *Bucchi* and *Trench* [24.53]. We have now *plural* science and *plural* public (diversity and fragmentation of publics); the new mediations by digital media have produced the *crisis of mediators* (losing filters that guarantee the quality of information, the actual information overload demands new definitions of quality and standards for assessment), and the traditional sequence of the communicative process from the specialist discussion/didactic exposition to the public communication (*popularization)* of *ready-made science* are changing towards *engagements* in *science-in-the-making* processes. Indeed, the users of scientific information increasingly have access to science in its making and highly controversial debates among specialists. Strong forms of engagements include *participatory democracy* and *participatory communication* that are connected to citizen science and open science [24.78]. *Rethinking science* [24.50] goes hand in hand with *reinventing discovery* [24.78]. Last but not least, the *cultural contexts*, which expresses models, and visions of knowledge translation and transfer is becoming more and more important. Finally, public communication has become a global enterprise with some overall features and specific regional characterizations [24.53].

All these changes affect the assessment of research. The changes described in the upper part of Table 24.2 have different consequences as reported in the bottom section of the table. In particular, they increased the *complexity of the research assessment and the need for standards*. Evaluating research and its impacts is a real complex task. *Daraio* [24.2] observed that perhaps the key problem is that research performance is not fully quantifiable. Hence, research assessment has to deal with *nonfully* quantifiable concepts. This has a consequence on the measurement of research productivity. It has to be carried out in a careful way, taking into account an appropriate representation of the production process, which identifies what are the boundaries of what is measured and describes all the main assumptions behind the representation.

Daraio and *Glänzel* [24.76] showed that the complexity of research systems requires a *continuous information exchange*. This process is due to the communication and interaction process among all actors and agencies involved in the production, processing and application of knowledge. All data entries, all processing, development and application of data relevant for research, and technology and innovation have their own rules and standards. *Daraio* and *Glänzel* [24.76] identify some elementary *rules of interferences* expressed in terms of data definition and standard setting in the process of data integration for different purposes, including process monitoring, input–output monitoring, ex-ante and ex-post evaluation.

In this chapter we maintain that *standards* and *harmonization* should be applied as well to the methodological dimension (or aspects) of the research assessment process. Although this is difficult due to the different and changing purposes of the assessment, we propose here to *standardize the methodological steps* of the econometric measurement of productivity and efficiency according to a general framework that will be introduced in the next subsection.

24.3.2 The Need for a Framework

The assessment of research requires a *systemic* approach in which research activities are considered together with education and innovation activities. In *Daraio* [24.2] we showed that the formulation of models of metrics is necessary to assess the meaning, validity and robustness of metrics. Metrics are defined as indicators calculated and used in a research assessment.

It was observed that developing models is important for *learning* about the explicit consequences of assumptions, testing the assumptions, and highlighting relevant relations; as well as for *improving*, by documenting/verifying the assumptions, systematizing the problem and the evaluation/choice done, and making explicit the dependence of the choice to the scenario. Moreover, there are several *drawbacks* in modeling, which have to be taken into account. The main pitfalls relate to the targets that are not quantifiable; the complexity, uncertainty and changeability of the environment in which the system works, to the limits in the decision context, and, last but not least, to the intrinsic complexity of calculation of the objective of the analysis.

A Summary

Daraio [24.2] proposes the framework illustrated in Fig. 24.1 as a reference to develop models of metrics. The results presented in this section have previously been published in [24.3].

This framework is based on three dimensions:

- *Theory*, broadly speaking, identifies the conceptual content of the analysis, answering the question *what* is the domain of interest, and delineating the perimeter of the investigation.
- *Methodology* generally refers to *how* the investigation is handled, what are the kinds of tools that can be applied to the domain of interest, tools which represent the means by which the analyses are carried out.
- *Data*, largely, and roughly, are instances coming from the domain of interest, and represent the means by which the analyses are carried out.

Each dimension is composed of three main building blocks and identifies three operational factors for implementation purposes. The main building blocks of theory are:

1. Education
2. Research
3. Innovation.

The main building blocks of methodology are:

1. Efficiency
2. Effectiveness
3. Impact.

The main building blocks of data are:

1. Availability
2. Interoperability
3. Unit-free property.

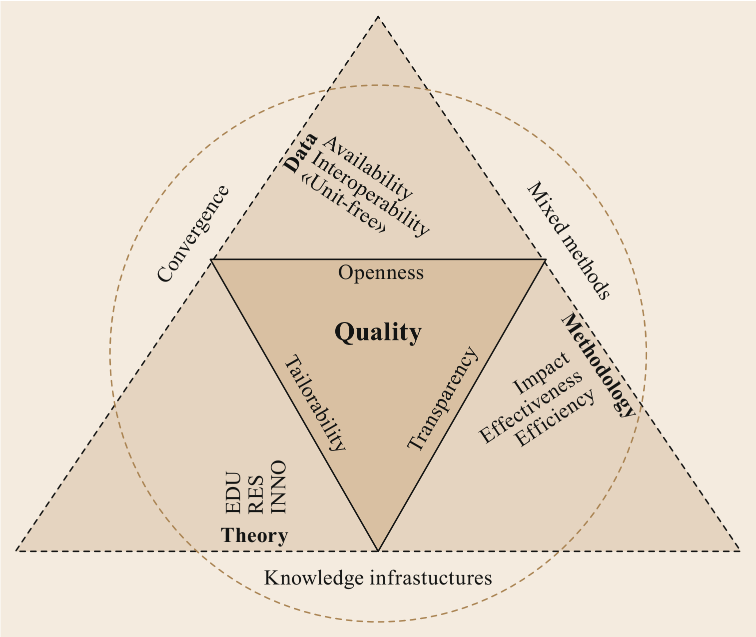


Fig. 24.1 A systemic framework for the development of research assessment models (after [24.2])

See Table 24.3 for their definitions.

The methodological dimension of this framework will be discussed in Sect. 24.6.

Daraio [24.2] asserts that the ability of developing (and afterwards understanding and using effectively) models for the assessment of research is linked and depends, among other factors, on the degree or depth of the *conceptualization* (formulation of the content of the general ideas and of the most important details) and *formalization* (refers to a defined structure), in an unambiguous way, of the underlying idea of quality. *Quality*, here, is intended as *fitness for use*.

The level of conceptualization and formalization of quality, however, is neither objective nor unique. It depends on the purposes and the subject or unit of the analysis (e. g., scholars, groups, institutions, up to meso or macro aggregated units, as regional or national entities) and it relates, in the end, to the specific evaluation problem under investigation.

In the next two subsections we report two examples of application of the proposed framework.

An Application of the Framework to the ANVUR's Activities

In a period of budgetary restrictions, policymakers need timely and inexpensive answers to their questions. While our framework highlights the need to invest in knowledge infrastructures as an investment for the development of a new generation of assessment models, it offers at the same time a pragmatic scheme to identify priorities in policy actions.

Coherent with our framework is the view of priority setting as a problem in system design, which is [24.80]:

> Best understood as a systemic process, with outcomes determined by the incentives and interrelationships of choice rather than by ex-ante calculation.

To show an example, we use the framework illustrated in Fig. 24.1 to frame the activities carried out by the Italian National Agency for the Evaluation of Universities and Research Centers (ANVUR). On the reform of the Italian system, see e. g., [24.81], while for some comparative analyses of the Italian system with France and UK respectively, see [24.82, 83].

In Fig. 24.2 we illustrate the main activities carried out by ANVUR so far, where Autovalutazione, Valutazione periodica, Accreditamento (AVA) stands for the evaluation of teaching, Valutazione della Qualità della Ricerca (VQR) for the evaluation of the quality of the research activity and Third Mission stands for the assessment of third mission activities of universities and research centers. ANPReS (Anagrafe nominativa dei professori ordinari e associati e dei ricercatori e delle pubblicazioni scientifiche) is an acronym that indicates the registry (Anagrafe) of the Italian scholars, foreseen in the current legislation but not yet implemented. For that reason, Fig. 24.2 reports it as a wide circle. Open Researcher and Contributor ID (ORCID) indicates the measure introduced by law to require each academic staff member to obtain an ORCID code, a successful

Table 24.3 Definition of the components of the framework [24.3, Table 1]

Dimension and component	***Definition***
1. *Theory*	Identifies the conceptual content of the analysis, answering the question *what is the domain of interest* and delineating the boundary of the investigation.
Education Research Innovation	These are the main conceptual blocks of theory. Their interrelations and their complementarities should be considered in a systematic way when assessing research.
2. *Methodology*	Identifies the range of methods, techniques and approaches that are relevant for the evaluation of research. It answers the question *how* the investigation is handled.
Efficiency Effectiveness Impact	These are the subjects of the assessment. They go from the *output* (baseline) that is the result of the transformation of inputs in outputs, to *efficiency*, which relates output to inputs with respect to an estimated efficient frontier, to *effectiveness*, which considers inputs, output and accounts for the aims of the activity, while *impact* refers to all contributions of research outside academia.
3. *Data*	Data are a relevant dimension often neglected in model building. Data have a problematic definition because this depends on their use not on inherent characteristics of the data [24.79, p. 74]. Data are instances coming from the domain of interest and represent the means by which the analyses are carried out.
Availability Interoperability *Unit-free* property	Refers to the usability of data, alternatives and choices that affect the data. This is the way in which heterogeneous data systems are able to exchange information in a meaningful way. Need of consistent and coherent observations across different levels of analyses.
Quality	*Fitness for purpose*. It is the overarching concept of the framework. It is also an attribute of the different dimensions of the framework.
Implementation factors	
Tailorability	Adaptability to the features of the problem at hand.
Transparency	Description of the choices made and underlying hypothesis masked in the proposed/selected theory/method/data combination
Openness	Accessibility to the main elements of the modeling.
Enabling conditions	
Convergence	Evolution of the transdisciplinary approach, which allows for overcoming the traditional paradigms, increasing the dimensional space of thinking.
Mixed methods	Intelligent combination of qualitative and quantitative approaches.
Knowledge infrastructures	Networks of people that interact with artifacts, tools and data infrastructures.

measure for the standardization and calculation of bibliometric indicators foreseen in the assessment exercise. Although rough and approximated, Fig. 24.2 immediately allows us to identify and indicate the priorities in the next planning of the activities, which are illustrated in Fig. 24.3 first along the data dimension, then along the methodology dimension (moving from the output towards efficiency, effectiveness and impact) and finally along the theory dimension.

A Doubly Conditional Performance Evaluation Model

This section is largely based on *Daraio* [24.4, 84]. We have seen that research performance is an articulated and complex concept [24.43] because the output of the research activity has some features that include complexity, uncertainty and indeterminacy.

Daraio [24.4] proposes a *doubly conditional performance evaluation model* as a *democratic evaluation tool* for *value creation* in a learning and participatory environment. It is considered as a revisited version of the Ricardo's approach of *comparative advantages* but in the context of a broader framework including theory, methodology and data dimensions. It is *doubly conditional* because the evaluation is conditioned twice: on the information *we have* and on the information *we do not have*. This model further operationalizes the doubly conditional performance assessment, identifying two kinds of conditions. The *internal conditions* include actors, processes, time, and results. These factors allow us to compare comparable entities, and to set appropriate reference sets. This is named as *internal conditioning* or normalization. There are in addition factors which account for the heterogeneity of the entity assessed, including the context, such as time frame, potential heterogeneity factors—other factors that may affect the performance not directly observed, standards (criteria or rules), and understanding incentives, actions and consequences that are named as *external conditioning* or contextualization. This model of performance evaluation, being based on a framework that allows us to describe and discuss the main assumptions of the

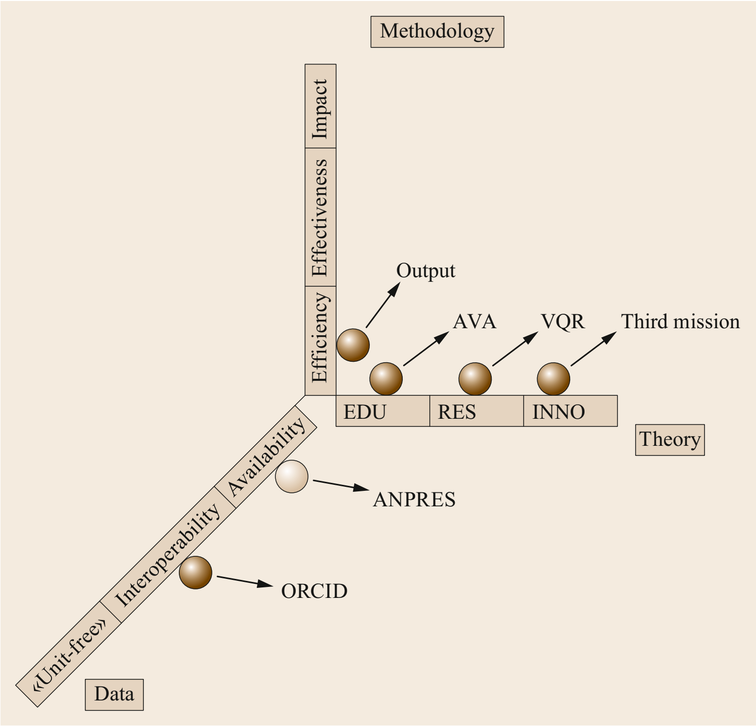


Fig. 24.2 A schematic map of the ANVUR's activities based on the framework illustrated in Fig. 24.1

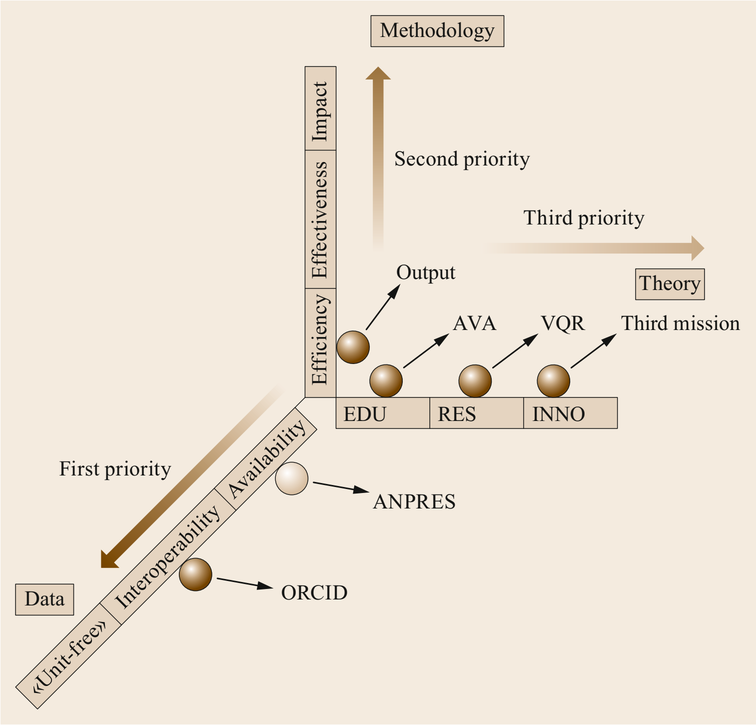


Fig. 24.3 Suggested priorities in the ANVUR's activities mapped in Fig. 24.2

model of the assessment, reduces the measurement of performance to a matter of appropriate normalization and contextualization.

This model may be helpful to *identify* the *components* of the analysis (in terms of theory-method-data characterization) that are *excluded* (what remains outside) in the specific context of the evaluation. It also helps to *interpret* the measure (or metrics or indicators) of research assessment calculated, and generates a *residual*, i. e., the part of the observed reality that is not accounted for—and thus *remains*, after the consideration of the relevant information is included.

The indicators empirically calculated can be interpreted as the *residual* or *our ignorance* on the phenomenon and enable one to identify the neglected aspects of the analysis carried out. This description of the model uses the term *residual* that is normally used in statistical analysis. Its use aims to underline a part of the observed reality that is not accounted for, but does not mean that this part is necessarily quantitative. The neglected component can be useful for suggesting alternative or additional dimensions of research assessment.

An underlying idea of this doubly conditional performance evaluation model is that *for each subject under assessment a dimension of performance can be found along which the evaluated entity can, in a relative way, outperform itself*. These are relevant factors among others for the development of a *participatory learning* role of the subjects involved in the research assessment exercise. A further discussion on the components of this performance evaluation framework will be presented in Sect. 24.7.

24.4 Economics and Econometrics in the Current Time

After the previous section, to understand the methodological challenges of economics and assessment, we need to understand the current evolution in the economics, econometrics and in the economic fields that may be related to research assessment.

Table 24.4 presents the enrichment of the definition of econometrics that took place over time. Starting from these definitions, we can try to understand what econometrics can offer to the evaluation of research. Econometric models are quantitative models that have been developed to model data and economic problems. As noted by *Baumol* and *Blinder* [24.88], economics is:

> A broad-ranging discipline, both in the questions it asks and the methods it uses to seek answers. Many of the world's most pressing problems are economic in nature.

Table 24.4 Definition of econometrics over time and other technical concepts. The section of the table on *other definitions* is taken from [24.84]

Definition of econometrics	Reference
Econometrics is the application of mathematics and statistical methods to the analysis of economic data	*Frisch* [24.85, p. 95]
Econometrics is the quantitative analysis of actual economic phenomena based on the concurrent development of theory and observation, related by appropriate methods of inference. The traditional econometric problems include: • Test of the optimizing behavior: parametric and nonparametric tests • Structural models and reduced form models • Estimation of technological relationships, more complex technologies • Choice of functional form • Estimation and aggregation • Hypothesis testing • Specification problems in econometrics.	*Samuelson* et al. [24.86, p. 142]
As a unified discipline, econometrics is still relatively young and has been transforming and expanding very rapidly. Major advances have taken place in the analysis of cross-sectional data by means of semiparametric and nonparametric techniques. Heterogeneity of economic relations across individuals, firms and industries is increasingly acknowledged and attempts have been made to take it into account either by integrating out its effects or by modeling the sources of heterogeneity when suitable panel data exist. The counterfactual considerations that underlie policy analysis and treatment valuation have been given a more satisfactory foundation. New time-series econometric techniques have been developed and employed extensively in the areas of macroeconometrics and finance. Nonlinear econometric techniques are used increasingly in the analysis of cross-section and time-series observations. Applications of Bayesian techniques to econometric problems have been promoted largely by advances in computer power and computational techniques. The use of Bayesian techniques has in turn provided the investigators with a unifying framework where the tasks of forecasting, decision making, model evaluation and learning can be considered as parts of the same interactive and iterative process, thus providing a basis for *real-time econometrics*	*Geweke* et al. [24.87]

Table 24.4 (continued)

Other definitions	Reference
Nonparametric methods	
Nonparametric statistical approaches are also named as *distribution-free* techniques because they do not require the specification of a functional form for the distributions of the variables analyzed and/or estimated. Nonparametric methods are often associated with techniques and tools that do not make numerous or stringent assumptions about the model underlying the empirical analysis. For this reason, compared to models underlying parametric methods, which rely on the specification and or assumptions about the functional form of the involved variables, the conclusions based on nonparametric inference are more general. Nonparametric tools are valid under less restrictive assumptions than classical parametric statistical tools. Frequently, nonparametric methods are associated with bootstrap and Monte Carlo techniques.	*Gibbons* and *Chakraborti* [24.89]
Bootstrap	
The bootstrap is a data-based simulation method for statistical inference. The essence of the bootstrap idea is to approximate the sampling distributions of interest by simulating (or mimicking) the data generating process (DGP), i. e., the statistical model of interest. Then, this sampling distribution is used to draw inference, which is used to calculate bias, confidence intervals and so on, on the quantity of interest. The aim of the bootstrap is to provide an approximation of the sampling distribution, which can be easily obtained by using Monte Carlo approximations.	*Efron* and *Tibshirani* [24.90]
Monte Carlo techniques	
Monte Carlo techniques deal with random experiments run on a computer. The main underlying idea is to repeat the experiment a large number of times to obtain many quantities of interest exploiting the law of large numbers and other methods of statistical inference. Monte Carlo techniques are used for sampling, estimation and optimization purposes for a variety of reasons, among which they are easy and efficient and provide insights into the randomness of the selected model. Monte Carlo techniques also permit us to have precise information on the accuracy or the efficiency of a given Monte Carlo estimator/algorithm, thanks to the mathematical and statistical justification underlying them.	*Kroese* et al. [24.91]

24.5 What We Could Learn from Economics and Management

We have seen in Sect. 24.2.2 that productivity measurement is a part of the research evaluation. The assessment process can be distinguished, for practical reasons, into three phases: design (ex-ante), implementation (in itinere) and conclusion (ex-post). In the planning phase, the adequacy of available resources with respect to the objectives to be achieved is analysed. In this phase, the assessment concerns the possibility of achieving the objectives and the necessary financial, human and time resources. The assessment *in itinere*, also called *formative evaluation*, establishes what works and what can be improved in the ongoing process. The ex-post assessment, also called *summative evaluation*, is carried out at the end of the process and concerns the obtained results, the outcomes [24.10].

As we have seen in the previous section, economics, in sum, deals with the allocation of scarce resources. It is connected with management that is related to *making human resources productive*, according to *Drucker* and *Maciariello* [24.92, p. xxvi].

Recent trends in management go from the Porter's competitive strategic environment [24.93–95] five forces (entry, internal rivalry, buyer power, supplier power, substitutes and complements) to the competition on analytics [24.96] to *beyond competitive advantages* [24.97] in terms of *value creation*.

Now let us see what aspects of the economic theory and management methods can be useful for the research evaluation process.

24.5.1 Strands of Economic Research

The main strands of economic research that can be useful, without claiming to be exhaustive, are:

1. *Microeconomics and economics of production* for the modeling of the production process and assessment of productivity
2. *Strategic analysis* for the management of the research assessment
3. *Cost–benefit analysis* about the research assessment process
4. *Measurement of the performance in the business sector* to find out, mutatis mutandis, similarities trends and lessons that can be learned
5. *Knowledge management and intellectual capital*, about the measurement of research productivity because research is connected to knowledge and intellectual capital.

Economics of Production and Economics of Science

The typical areas of economics concerned are the *economics of production* and the *economics of science*. They were introduced and discussed in Sect. 24.2.4. We will come back to this topic in Sects. 24.6 and 24.7.

Strategic Analysis

According to Chandler:

> Strategy is the determination of the basic long-term goals and objectives of an enterprise, and the adoption of courses of action and the allocation of resources necessary for carrying out these goals.

Strategy [24.98] is important for the management of the assessment process and for developing economic models that include different perspectives (mathematical perspective: to discover the logic of choices; psychological perspective: motivations and behaviors; organizational perspective; political science perspective; and so on). To this purpose, (economic) models must carefully identify:

Decision makers Who are they? Are their decisions fixed?,
Goals profit maximization, nonpecuniary interest. . . ,
Choices actions, strategic variables, time horizon and
Relationship between choices and outcomes what the mechanism is to translate specific decisions into specific outcomes; complication due to uncertainty related to taste, technology, choices of other decision makers . . .

A classical framework for analyzing strategy in industrial organizations [24.98] is based on four broad classes of issues:

1. *Boundary* of the institution (what should the institution do, how large should it be, and what business should it be in)
2. *Market and competitive analysis* (nature of the markets in which the institution competes and nature of competitive interactions among institutions in those markets)
3. *Position* and *dynamics* (position of the institution to compete, basis of its competitive advantage, how to adjust it over time)
4. *Internal organization* (organization of institution's internal structure and system).

We observe that some recent developments in econometrics and the availability of new data and information show converging movements with the economics of strategy summarized above. A positive definition of data includes [24.99]:

> Facts and statistics collected together for reference or analysis; data as representations, reinterpretable representation of information in a formalized manner, suitable for communication, interpretation, or processing, up to data as infrastructure.

In particular, the study of *heuristics* in decision making introduced by *Tversky* and *Kahneman* [24.100] extended *Simon*'s research on human-bounded rationality in problem solving [24.101–103] which led to the *satisficing* situation where people seek solutions or accept choices or judgments that are *good enough* for their purposes. The discussion on heuristics in human decision-making and of their inherent biases is extended in *Kahneman* [24.104] who, building on earlier contributions, describes two different ways of thinking, a *fast system*, characterized by fast, automatic, frequent, emotional, stereotypic subconscious and a *slow system*, characterized by slow, effortful, infrequent, logical, calculating conscious. On the basis of heuristics, *Kahneman* [24.104] asserts that the fast system involves the association of new information with existing patterns instead of building new patterns for each new event. These recent developments in behavioral economics and decision making could be further explored in combination with recently developed statistical and machine learning approaches [24.105–108]. Indeed, machine learning techniques, lying at the intersection of computer science and statistics, is at the core of artificial intelligence and data science, and is showing increasing potentialities [24.109, 110].

Cost–Benefit Analysis

The four main valuations of a cost–benefit analysis (CBA) have been identified in [24.111, p. 3]:

> 1 The relative valuation of costs and benefits at the time when they occur. 2 The relative valuation of costs and benefits occurring at various points in time: the problem of time preferences and the opportunity cost of capital. 3 The valuation of risky outcomes. 4 The valuation of costs and benefits accruing to people with different incomes.

The same authors [24.111, p. 4] describe the process of a cost–benefit analysis as organized into two phases:

> (a) *Value the costs and benefits* in each year of the project, (b) *Obtain an aggregate 'present value'* of the project by 'discounting' costs and benefits in future years to make them commensurate

> with present costs and benefits, and then adding them up. At each stage, the evaluation differs from commercial project appraisal because (i) costs and benefits to all members of society are included and not only the monetary expenditures and receipts of the responsible agency, and (ii) the social discount rate may differ from the private discount rate.

For more information, see e. g., *Boardman* et al. [24.112].

Weimer [24.113], in the introduction to the special issue on *Cost–Benefit Analysis and Public Policy* states that:

> CBA holds a prominent, but controversial, place among the techniques of public policy analysis. At one extreme, some economists view CBA as synonymous with good policy analysis. At the other extreme, a diverse group of political philosophers attack it as a technocratic undercutting of democratic values, a utilitarian threat to individual rights, or a crass debasing of public discourse. Yet, CBA is neither panacea nor fatal poison. Though often impractical to implement and rarely fully appropriate as a formal decision rule, it provides policy analysts with insights for organizing their thinking about the goal of efficiency and specific techniques to help guide the measurement and valuation of the impacts of policies in terms of the resources they require and the effects they produce.

Assessment and Management of Performance in Business

The following section presents results previously published in *Daraio* [24.4]. *Daraio* [24.4, 84] noted that in the evaluation of research, there has been an evolution from a traditional performance evaluation, based on traditional bibliometric indicators of number of publications and citations, towards a multidimensional performance model, which includes alternative and impacts metrics. According to *Ghalayini* and *Noble* [24.114], a similar pattern has been observed in business performance measurement. In the first phase (from 1880s to 1980s) the emphasis was on financial measures such as profit, return on investment, and productivity. The second phase started in the late 1980s as a result of changes in the world market. To regain a competitive edge, companies shifted their strategic priorities from low-cost production to quality and flexibility, but also implemented new philosophies of production management (computer-integrated manufacturing, just in time, total quality management and so on). These changes have revealed that there is a need to develop new performance measurement systems. As the *Performance Measurement Manifesto* [24.115, p. 131] states:

> At the heart of this revolution lies a radical decision: to shift from treating financial figures as the foundation for performance measurement to treating them as one among a broader set of measures.

In the Performance Manifesto the revolution is all about combining information systems and human resources (called a culture shock).

In the business literature [24.103, 116] there has been an evolution of the performance evaluation from the classical *shareholder* value approach (based on the maximization of profits) to the balanced scorecard approach [24.117], which extends the performance measurements to four dimensions (4-D). In addition to the classical economic perspective, they include also the other perspectives of internal process, of customers, and finally of the growth and learning. Recent examples of multidimensional performance evaluation models are the multistakeholder model by *Atkinson* et al. [24.118] and the performance prism model by *Neely* et al. [24.119]. The most recent trends in this sector are the inclusion of indicators of sustainability and the development of social corporate responsibility awareness.

In the assessment of research the Leiden manifesto [24.67, 68] can play, mutatis mutandis, the same role of the performance manifesto for the business sector.

Daraio [24.4] showed the extension of the principles of performance measurement of the business sector to the public sector, in the so-called new public management [24.120–122]. Recent trends, described at length in *Van Dooren* et al. [24.123] include a trend towards the *democratization* [24.124–127] linked to the need for a *citizen-driven* performance measurement.

From the *management of performance* in the public sector literature we can learn about the use of performance information in policy and management practice. *Van Dooren* et al. [24.123, p. 119, Table 6.1] for instance, list 42 potential uses of performance information including changing work processes, strategic planning, communication with the public to build trust, reporting and monitoring, accountability, clarifying objectives, cost–benefit analysis, staff motivation, nonmonetary incentives and so on. They consider three main uses of performance information. These are to learn, to steer and control and to give account. To these main uses, they connect the manifestation of functional or dysfunctional *effects*. As noted by *Van Dooren* et al. [24.123, pp. 183/184]:

> The literature on the dysfunctions of performance measurement is much richer than that on the

functions. It includes: i) manipulation of the measurement process (over and underrepresentation, failing measurement, *mushrooming* – increasing numbers which inflates their values of indicator sets, *polluted* performance information, unjustifiable aggregation or disaggregation of data, *misrepresentation* ranging from creative accounting to fraud, misinterpretation – incorrect inference about performance and ii) manipulation of the output (measure fixation – oversupply of products, loss of quality – myopia – short-termism, suboptimization, *cream skimming* or cherry picking, complacency – risk aversion and less excellence – organizational paralysis), as well as unintended consequences.

The outcome of these behavioral effects is the well-known performance paradox [24.128], which states that organizations adapt their behavior to reach the targets. This means that indicators, over time, lose their discriminatory power. Individual behavior is comprised of a more complex configuration of institutions that determine the conditionality of functions and dysfunctions, including but not limited to incentive structure, and other structural and cultural characteristics [24.123, p. 192]. *Van Dooren* et al. [24.123, p. 198] observe that:

> The occurrence of effects depends on the way performance information are used, besides some general cultural and institutional variables. Dysfunctional effects can be tackled by taking away the motive or the opportunity to behave dysfunctionally.

In discussing about the challenges and the future of performance management, *Van Dooren* et al. [24.123, p. 216] state that "Performance management systems need to be able to deal with the complexity in the environment". Performance management systems should facilitate learning and should be based on better quality data, more ownership, stronger leadership, integration, training and expectations management.

Knowledge Management and Intellectual Capital

The measurement of intellectual capital [24.129] is a prominent research area in knowledge management [24.130–132]. Measuring and managing the intellectual capital of *communities* is an original area of research that has the potential to *change how public sector planning and development is done* according to *Bounfour* and *Edvinsson* [24.133]. Intangibles and intellectual capital are important to the private sector and they are also important to the productivity and competitiveness of the public sector [24.134–136]. Within this area, it could be interesting to explore the principles of intellectual capital efficiency [24.137] and the value added intellectual coefficient [24.138] and their applicability within the assessment of research.

24.6 Methodological Challenges in the Assessment of Productivity/Efficiency of Research

In Sect. 24.3, following *Bonaccorsi* and *Daraio* [24.1], we stated that in contrast to a standard production activity, the production process of the research activity is characterized by several conceptual and measurement problems that affect any definitional elements of the productivity, namely the inputs, the outputs and their functional relation. The changes that occurred in the production and communication of science affect the representation of the production process and challenge the measurement of the productivity of research. In the following, we describe the main challenges in economic modeling, which may also be potential sources of *ambiguities* for the measurement of productivity/efficiency.

24.6.1 Changes and Challenges in Modeling

The assessment of productivity and efficiency is subject to different modeling changes and challenges. Some of them are summarized in Table 24.5.

The modeling challenges of Table 24.5 expand the problems already highlighted by *Bonaccorsi* and *Daraio* [24.1] that now are amplified by the need of clearly identifying and modeling the representation of the production process behind the measures proposed. There is a need to identify clearly the *boundary* of the activities to be measured. Now the representation of the research activity process is complicated by the evolution summarized in Table 24.5. The *multidimensional* nature of the research activity, its multiple inputs–multiple outputs, the nondeterministic relation between inputs and outputs that are heterogeneous and sometimes incommensurable, and the complex lag structure of the output in the current time are affected as in Table 24.5 and pose a series of conceptual and methodological problems. They will be analysed after the description of the advent of networks in economics, another challenge posed to the econometric assessment of productivity and efficiency.

Table 24.5 Changes and challenges of modeling

Changes (adapted from [24.139])	
From	**To**
Well controllable models of the past	Complex models of today
Weakly connected or independent system components	Strongly connected or independent model component
Dominated by the model components	Dominated by their interactions
Simple model behavior	Complex model behavior
Sum of properties of individual components characterizes model behavior	Emergent collective behavior implies new and often unexpected model behavior
Conventional wisdom works well	Counterintuitive behavior, extreme events are common
Well predictable and controllable in top-down fashion	Less predictable, management by setting rules for bottom-up self-organization
Challenges	
Challenges are in	**Source**
The formulation of the problem, selection and evaluation problems	*Doornik* and *Hendry* [24.140]
Theoretical selection of the model and appropriate estimation technique. Statistical model selection. Implementation of automated general-to-specific model selection (so-called *autometrics*)	*Doornik* and *Hendry* [24.140]
Significance versus spuriousness	*Ekbia* et al. [24.141]
Causation versus correlation	*Ekbia* et al. [24.141]

24.6.2 The Advent of Networks in Economics

The economic systems are more and more conceived *complex* ecosystem whose *operating models* have to be developed considering the interplay of many dimensions.

In economics we are witnessing the development and expansion of networks [24.142–144] up to the point that *Kirman* [24.145] posed the question about networks as *a potential paradigm shift for economics*.

As observed by *Mandell* and *Keast* [24.146], networks have been assessed on the basis of traditional measures while ignoring the importance of process variables and their impact on outcomes in networks. Furthermore, networks can be complex arrangements, operating within and across layers of interaction with diverse member expectations and goals. Within this framework, different types of evaluation processes are needed to incorporate the complex and unique characteristics of networks.

In this context, there is an interest in looking for new models of production process representation. This research could be connected to recent developments in econometrics of information [24.147, 148] statistical inference and machine learning [24.107, 108]. Within this framework, new *reuses* of concepts already known in the past may be helpful.

Let us consider for example, *Georgescu-Roegen*'s [24.149–152] model of *flows and funds*. It allows an analytical representation of the *organization* of the production process that goes beyond the relationship between the inputs and the outputs, which is typical of production functions and input–output analyses. It allows us to include the organization and time dimension of production processes [24.153]. The *flows and funds* model may be connected to the neo-Schumpeterian interpretative framework of production of new processes by means of creation and diffusion of knowledge [24.154] in which there is an *interplay* between capabilities, transactions and scale and scope to explain the boundary and the competitiveness of the analysed units [24.155]. *Fioretti* [24.156] operationalizes the *organizational aspects* of production (typical of the flows and funds model) by linking them to recent neural networks approaches.

There is a rich literature on the nonparametric estimation of efficiency based on networks, called network DEA. *Kao* [24.157] offers an overview of these models, which all aim to asses the performance of complex systems. *Daraio* et al. [24.158] propose a general framework that embraces the Georgescu-Roegen model with information theory and statistics of complex systems and thus allows us to estimate the interdependencies between productivity networks. They propose a pseudolikelihood approach to infer in a Bayesian context the topology of the network structure analyzing the interdependencies among productivity networks. This is a promising line of research to further explore.

24.6.3 Conceptual and Methodological Ambiguities

In the following, the line of argumentation as previously published in *Daraio* [24.2] is presented. Every model is subject to *uncertainty*, stemming from uncertainty in the input information fed to the model, e. g., from lack of data or understanding of the governing relations, the fact that the model is a *simplification* of reality involving important elements of *subjectivity*. It is therefore important to acknowledge and quantify the uncertainty in the results of models to ensure a transparent and high-quality usage of them. We will come back to modeling in Sect. 24.7.

Conceptual ambiguities are relative to the definition of the components: what do we measure and how do we measure it? Methodology, in this setting as in Table 24.3, identifies the range of methods (techniques or approaches) that are relevant for the evaluation of research. The methodological dimension should handle how to evaluate what, providing an appropriate account of reliability and robustness and uncertainty.

The discussion on methodology relates two general interconnected questions that are *what to assess* and *how to assess* [24.2]. These questions, in turn, are related to the organization of the assessment tasks and strategies (including *priorities setting*) and to the communication of the assessment results.

Daraio [24.2] distinguishes the *subject* of the assessment from the means of the assessment. The *subject* of the assessment is the thing that is being considered of the assessment (*what to assess*). The subject of the assessment is identified in outputs, efficiency, effectiveness and impact. The *means* of the assessment (*how to assess*) can be *qualitative* (including peer-review and case studies), *quantitative* (including econometric approaches and tools from the physics of complex systems) and *combined* (quantitative–qualitative) approaches, including the so-called informed peer-review. Evidently, the means should be chosen in accordance with the subject of the assessment.

The organization and the communication aspects of the evaluation, however, fall within the sphere of policy and governance. The framework illustrated in Fig. 24.1 proposes three building blocks for methodology: efficiency, effectiveness and impact, considering the outputs as a kind of baseline or step zero in the analysis. Table 24.3 provides a definition of these concepts. Moving from efficiency to effectiveness, including quality indicators to assess effectiveness instead of efficiency, is another important step, which may go further, up to including impacts.

Classical methods of impact assessment [24.159] are challenged by [24.160, p. 22]:

> [The] problem of evaluation [that] is that while the program's impact can truly be assessed only by comparing actual and counterfactual outcomes, the counterfactual is not observed. ... Finding an appropriate counterfactual constitutes the main challenge of an impact evaluation.

These classical methods appear inadequate to the checklist of *sensitivity auditing* [24.161–163], that is, a sensitivity analysis applied to the modeling phase.

Alongside the conceptual ambiguities, there are also methodological problems. See Table 24.6 for an outline.

24.6.4 Data Issues

The explanations outlined in this section are based on *Daraio* [24.2]. Data are a relevant dimension often neglected in model building.

The problems of data in econometrics have been analysed for many years in econometrics (the so-called *data constraints* described in [24.181, 182]).

Data have a problematic definition because it depends on their use not on inherent characteristics of the data [24.79, p. 74]. Their properties and their weaknesses affect both the modeling and the empirical results. The concepts of [24.79, p. 271] state:

> Big data, little data, and even no data remains poorly understood in the current big data era: Efforts to promote better data management, sharing, credit, and attribution are well intentioned, but stakeholders disagree on the starting points, the end goals, and the path in between. Lacking agreement on what entities are data, it remains difficult to establish policies for sharing releasing, deposing, crediting, attributing, citing, and sustaining access that can accommodate the diversity of data scholarship across domains. Sustaining access to data is a difficult and expensive endeavour.

Borgman [24.79, p. 287] states:

> Despite the overall lack of agreement, most scholars would like better means to manage whatever they do consider to be their data. Better management is likely to lead to more sustainable data and in turn to better means of discovering and sharing data. These, however, are expensive investments. Better access to data requires investments in knowledge infrastructures by research communities, funding agencies, universities, publishers, and other stakeholders.

We propose to characterize the data issues according to the main dimensions of data of our framework described in Sect. 24.3. They are availability, interoperability, *unit-free* property.

Table 24.6 Main features and typical problems of econometric approaches to assessing productivity

Reminder of basic notions	**Parametric versus nonparametric**	**Deterministic versus stochastic**
	In *parametric models* the frontier is a known mathematical function depending on some unknown parameters. Main advantages: the economic interpretation of the parameters and the statistical properties of the estimators. The main drawbacks are the choice of the function and the handling of multiple outputs cases. *Nonparametric models* do not assume any particular functional form for the frontier function. Main pros: robustness to model choice and easy handling of multiple inputs–outputs cases. Main cons: estimation of unknown functional (more difficult) and *curse of dimensionality* (that is typical of nonparametric estimators and means the need to have a large amount of data to avoid large variances and wide confidence intervals)	In *deterministic models* all the observations are assumed to belong to the production process with probability one. The main drawback is the influence of outliers. In *stochastic models* it is assumed that there may be noise in the data. The main drawback is the identification of noise from inefficiency.
	Production functions	**Production frontiers**
Object of the estimation	Average (representative) behavior, conditional expected value	Efficient frontier, best practice (envelope)
Methodological problems of most used approaches	**Methodological problems of parametric production functions**	**Methodological problems of nonparametric production frontiers**
	Identification: the fundamental issue of whether the parameters of interest in the model are estimable [24.164].	*Deterministic* nature. In this framework it is assumed that all deviations from the efficient frontier are owed to inefficiencies
	Misspecification concerns the problems and errors related to the assumptions made by the model. Empirically, misspecification errors are mainly related to the specification of explanatory variables, in particular, knowledge of which one of the variables to include and about the mathematical form of their inclusions	Difficult economic interpretation (due to the *lack of parameters*) of the production process in terms of, e. g., shape of the production function, elasticities, and so on
	A topic related to the previous one is the *exclusion of relevant variables* and the *inclusion* of *irrelevant* variables	*Exclusion of relevant variables* and the *inclusion* of *irrelevant* variables is an issue also for the frontier analysis
	Simultaneity in the relationship between variables could greatly affect the estimation of parameters creating a source of bias.	*Curse of dimensionality*. Shared by many nonparametric methods the curse of dimensionality means that to avoid large variances and wide confidence interval estimates a large quantity of data is needed
	Multicollinearity is the problem related to the existence of a linear dependence amongst the response or independent variables. The multicollinearity affects the problem of unidentifiability of the regression parameters.	Difficulty in making statistical inference, owing to its complex nature: nonparametric estimation in a space at $p+q$ dimensions (where p is the number of the inputs and q is the number of the outputs), based on very few assumptions.

Availability refers to general alternatives and choices that affect the data that have to be used, for instance (without being complete): sampling versus census, freely available versus controlled or undisclosed ones, data as consumption versus participation (see [24.141] for a critical discussion). Obviously, the minimal requirement for the elaboration of data refers to their availability in a usable way. This opens to the discussion on commercial versus publicly available (or open) data; institutionally provided data; and issues of privacy and confidentiality. *Interoperability* is the way in which heterogeneous data systems are able to communicate and exchange information in a meaningful way [24.183]. It is crucial for data integration of heterogeneous sources (see the discussion on continuity versus innovation in [24.141]). *Unit-free property* refers to the need to have consistent and coherent observations (instances of data) at different levels of analysis, to ensure robust empirical evidence of a given phenomenon. The unit-free property of data is somewhat interconnected to the possibility of multiscale modeling of the problem at hand. It clarifies the exigence of having data that are *independent* from the unit of analysis and hence can be used *coherently* in a multiscale model of the problem.

Table 24.6 (continued)

Proposed solutions and more general approaches	Proposed solutions and more general approaches	Proposed solutions and more general approaches
	Simultaneity could be controlled for by applying generalized methods of moments (GMM) [24.165]. GMM is a general statistical approach that subsumes different estimation methods of interest, such as least squares, maximum likelihood and instrumental variables (see [24.166] for an introduction to these methods).	The problem of *extremes or outliers* and *the curse of dimensionality* can be treated by applying *robust partial frontiers* (for an overview, see [24.13]). The order *m* frontiers for instance represent a more realistic benchmark. Instead of comparing the performance of each unit with the best performers, the benchmark is done against the expected value of an appropriate sample of *m* units, drawn randomly from the population. The method offers flexibility in choosing the level of robustness of the estimate, by varying the parameter *m*. Robust partial frontiers have the nice property of not being affected by the curse of dimensionality.
	Semiparametric methods [24.16, 167] may be applied to reduce the burden of the assumptions required for estimation and inference by parametric approaches	The problem of handling noise in this context is owed to the model not being identified unless some restrictions are assumed. See, e. g., *Aigner* et al. [24.168] for approaches that assume a parametric function for the frontier; or *Kneip* and *Simar* [24.169] for the case of panel data. More general results for handling noise in nonparametric frontier models have been introduced in [24.170, 171] that attempt to address the identification issue (i. e., distinguishing between what is noise and what is inefficiency) by assuming the less possible structure. *Kumbhakar* et al. [24.172], *Simar* and *Zelenyuk* [24.173], *Park* et al. [24.174] propose local likelihood approaches. Recently, *Florens* et al. [24.175] relaxed the normality of the error term, assuming only a symmetric error with unknown variance.
	Production functions may be estimated by nonparametric regression techniques that offer the following advantages [24.176]: 1. Allow us to estimate functions of greatest complexity 2. Offer the possibility to make predictions without relying on specific parameters 3. Identify spurious observations and outlying points 4. Represent flexible methods for data interpolation and missing values imputation.	The problem of lack of parameters for economic interpretation may be overcome by *parametric approximation of nonparametric and robust frontiers* that have been introduced by *Florens* and *Simar* [24.177] and extended to the multivariate input–output case in *Daraio* and *Simar* [24.13]. These techniques have not been applied so much up to now, but have a great potential.
		Recent development in statistical theory for efficiency estimators provided *statistical inference in nonparametric frontier models*. In particular, the development of new central limit theorems opened the door to testing hypotheses about the structure of production models (for an overview, see *Simar* and *Wilson* [24.178, 179]; an updated review can be found in *Mastromarco* et al. [24.180]).

The data problems and their impact on the measurement of research productivity and performance have been analysed both at the macro level [24.184, 185] and at the micro (institutional) level [24.186–188]. Most of the problems relate to the measurement of the inputs, and of the outputs to their combinations. Generally, being collected from different sources and almost independently, the combination of these data is affected by comparability and consistency problems. Recent experiences on the micro data related to European universities have shown that taking into account the different data quality aspects of the data [24.189], a data quality-aware usage of them it is still possible. Nonetheless, a lot of further work is required in this area.

24.6.5 The Implementation Problem

The implementation of the econometric models is a delicate activity. The problem of implementation consists of applying methods developed as *basic research* to concrete organizations/contexts. The main critical points are the *interaction* of method development with its useful application and that the implementation changes the organization. The identification of the *right* problem and the development of an *appropriate* model are then crucial determinants of success.

Our framework, introduced in Sect. 24.3, proposes three *implementation* factors:

1. Tailorability
2. Transparency
3. Openness.

(See Table 24.3 for their definitions.) According to *Daraio* [24.2] the more one is able to go to the deep, fine-grained end of the most atomic-level unit of analysis (i. e., the higher the level of tailorability), the higher the level of openness and transparency, and the better will be the conceptualization and formalization of *quality* within a model.

As far as the implementation of models of research assessment is concerned, *Daraio* [24.3]—to which the reader is referred to for a deeper discussion and a graphical illustration—has shown the interplay existing between the problem context, the abstraction/ontological commitments, and the social translations. These latter are in place in the assessment of research because research is a human activity.

24.7 Potential of Econometric Approaches and of Nonparametric Methods

24.7.1 Models for Research Assessment

Based on [24.2], it is argued here that a model is an abstract representation, which from a particular point of view and for a particular purpose represents an object or real phenomenon. The representation of reality is achieved through the *analogy* established between aspects of reality and aspects of the model. Econometric models are quantitative models: models in which the *analogy* with the real world takes place in two steps:

1. Quantification of objects, facts and phenomena in an appropriate way
2. Identification of the *relationships* existing between the previously identified objects, closest to the reality (that is the object of the model).

The practical use of a model depends on the different roles that the model can have (*all models are wrong, but some are useful* [24.190]) and from the different steps of the decisional process in which the model can be used.

Models may have four main roles:

1. Description
2. Interpretation
3. Forecasting
4. Intervention.

These roles may be correlated or not, depending on the objective of the analysis and the way the model is built.

To be successful the model has to take into account the specificities of the processes and systems under investigation. Behavior is free and finalized to given aims. History and evolution matter as the behavior of systems and processes change over time.

The *finalization* encourages a functional analysis of the systems: the external behavior of the systems may be explained by focusing the analysis on their aims and to their ways of interacting with the environment without entering into the details of the internal structures and organization (the organization becomes relevant only if it is a limit to pursue the objectives of the system).

The main difficulties which arise in modeling are:

1. The possibility that the targets are not quantifiable, or are multiple and conflicting; or that there are several decision makers with different interests
2. Complexity, uncertainty and changeability of the environment in which the controlled system works and thus the difficulty of predicting environmental stimuli, the consequences of certain actions, and responses to other decision makers in these actions
3. The limits (in particular of an organizational nature) with which the controlled system adapts to the directives of the decision maker
4. The intrinsic complexity of calculation of the optimum behavior.

The analyses of current changes in research and econometrics has lead us to consider the measurement of productivity/efficiency of research in a wider context.

The framework introduced in Sect. 24.3 and the evaluation model described in Sect. 24.3.2, *A Doubly Conditional Performance Evaluation Model* can be helpful in the development of models of research assessment.

24.7.2 Advanced Efficiency Methods

Nonparametric Methods

The evolution of econometric tools in recent years relates mainly to the development of nonparametric econometrics ([24.87], see definitions in Table 24.4). Nonparametric econometrics consist of the application of nonparametric techniques to analyze economic models and data. *Stock* [24.191] traces the development of econometric models from the traditional ones of the 1980s, mostly parametric, characterized by a linear functional form, to more recently developed nonparametric ones, thanks to the development of computer power and the advancements of mathematical and statistical research. *Stock* identifies one of the causes of the development of nonparametric models as dissatisfaction towards traditional parametric models that were not always a good approximation [24.191, pp. 84/85]:

> The past three decades have seen significant changes in the tools of econometrics, many motivated by a desire to minimize the effect of 'whimsical' assumptions on inference about the object of interest. By 'whimsical' I mean arbitrary assumptions that are subsidiary to the empirical purpose at hand, but which affect inference about the causal effect of interest. The new tools provide reliable inference without implausible subsidiary assumptions.

Today, nonparametric econometrics [24.16, 167, 192] is a well-established field both in theory and in practice.

We believe that nonparametric approaches are the most appropriate for research assessment for their generality. The success of the nonparametric approach is mainly due to the few assumptions required for specifying the data generating process (DGP). However, this approach presents also some limitations, namely the difficulty in carrying out statistical inferences, the curse of dimensionality, which are specific to the nonparametric estimators and the influence of extreme values and outliers, as outlined in Table 24.6.

As for higher education [24.3], the challenges determined by the advent of big data include the developments in scientific computing of simple techniques with a general applicability as the Monte Carlo tools; the availability of millions of data; and new administrative data available, which will allow us to overcome the traditional *curse of dimensionality* problem of nonparametric approaches, and will offer the possibility to extend and apply Bayesian inference and machine learning approaches [24.193]. Although these possibilities are mitigated by the dilemmas created by the big data [24.141] and the problems of data quality, data integration and model selection [24.140] are important open issues, and we see a great potential for nonparametric approaches in research assessment in the future. Big data need to be exploited [24.193–198]. However, taking full advantage of the potential of big data will require increasing sophistication in knowing what to do with the massive amounts of data that are now available [24.199]. This involves taking into account, depending on the field of application and of the context of data use, also the complementarities, the strengths and the pitfalls of the integration of big data with little data and not available and/or not usable data [24.200].

Evolution of Production Frontiers

The econometrics of production functions is different from that of production frontiers as the main objective of their analysis differs: production functions look at average behavior whilst production frontiers analyze the whole distribution, taking into account the best/worst behavior. Obviously, assessing the impact on the average performance is different from assessing the impact on the best/worst performance. Accounting for inequality and diversity is much more natural in a model based on best/worst performance frontiers than in a standard (average, representative) behavior. This is because in the former case the whole distribution is considered instead of only the central tendency. This distinction between *average* versus *frontier* is considered in recent theory of growth [24.201–203] and in the managerial literature [24.204].

All methods have advantages and disadvantages. As far as quantitative methods are concerned, different approaches, both parametric [24.205] and nonparametric [24.206–208] have been proposed, highlighting the changes required by the attempt to disentangle the impact of external heterogeneity factors on the efficient frontier from that on the distribution of inefficiency. This trend witnesses the need to move from the assessment of efficiency towards the assessment of impacts.

The traditional classification of frontier estimation methods in parametric versus nonparametric and deterministic versus stochastic is obsolete and no longer adequate to the changes and challenges summarized in Table 24.6. Table 24.7 shows some selected pioneering studies of the recent evolution of frontier approaches. *Daraio* and *Simar* [24.13] observed a general trend of

Table 24.7 Some precursors of the evolution in the frontier estimation (Source: [24.2, p. 20] with updates)

Approach	Main reference	Trend
Statistical approach to nonparametric frontier estimation	*Mastromarco* et al. [24.180], *Daraio* et al. [24.209]	Trend towards data-driven modeling
Models averaging in stochastic frontier estimation	*Parmeter* et al. [24.210]	Trend towards robustness of modeling
Using information about technologies, markets and behavior of institutions in productivity indices	*O'Donnell* [24.211]	Trend towards more comprehensive informational setup
From an implementation point of view, interactive benchmarking	*Bogetoft* [24.212]	Trend towards developing analytics for policy decision-making support
Moving from efficiency to effectiveness	*Simar* et al. [24.213], *Daraio* et al. [24.209], *Bădin* et al. [24.108, 206], *Daraio* and *Simar* [24.208]	Trend towards including (unobserved) heterogeneity, contextualization and estimating *quality* as unobserved heterogeneity factors and its effects, trend towards impact assessment

1. **Model development components:**
   a) *THEORY.* In the assessment of research have you considered all the systemic connections among education, research and innovation?
   b) *METHODOLOGY.* Have you identified the *subject* (what to assess) of the analysis and the *means* (how to assess–method) of it?

- ***Subject*** of the analysis may be:
  -*Output* of the assessment (baseline) result of a transfonnation process which uses inputs to produce products or services
  -*Productivity* partial or total factor productivity and *Efficiency* (productivity with respect to a reference)
  -*Effectiveness* considers inputs, outputs and account for the aims of the activity
  -*Impact* all conhibutions of research outside academia.
- ***Means*** of the assessment may be:
  -*Quantitative* approaches
  -*Qualitative* approaches
  -*Quali-quantitative* approaches.

   c) *DATA. 1) availability.* Have you assessed the usability of data?
   Consider: sampling vs census
   freely, controlled or undisclosed
   consumption vs participation
   commercial vs publicly available
   open, institutional provided
   privacy/confidentiality
   2) *interoperability.* Which level of interoperability do you have?
   3) *unif free property.* Are the data independent from the unit of analysis? What is the Ievel of their objectivity?

2. **Implementation factors:** tail or ability, openness and transparency. Assess the degree of the implementation factors (1 = very low, 2 = low, 3 = average, 4 = high, 5 = very high).
3. **Enabling conditions:** Convergence, mixed methods and knowledge infrastructure. Assess the presence of enabling conditioning (1 = not present, 2 = moderately present, 3 = present, 4 = high presence, 5 = very high presence).

Fig. 24.4 A preliminary checklist to develop econometric models for the assessment of research

convergence between parametric and nonparametric approaches, in that each one tries to overcome its limits, attempting to adopt the pros of the other approach. *Parmeter* and *Kumbhakar* [24.214] point to the great potential of nonparametric estimation in the stochastic frontier approach, while *Sickles* and *Zelenyuk* [24.215] summarize most recent developments in semiparametric models.

Now, it seems that we are going towards a multiple-mixed perspective (a mixture of parametric-nonparametric/deterministic-stochastic configurations). In this context, semiparametric approaches, including parametric approximations of nonparametric approaches may receive more and more interest from applied researchers. This evolution in the methodological development however is not always followed by the availability of software tools for empirical implementation. There are indeed many software options for running productivity and efficiency analysis (see a recent survey in [24.216]) although the quality of

Table 24.8 Doubly conditional performance evaluation model components and main questions for checklist development

Performance evaluation model component	Constitutive elements	Main question
Purpose of the assessment	Objectives, stakeholder and policy	*Why*
Level of analysis	Actors (micro level: scholars, organization; meso level: regional system; macro level: country)	*Who*
Object of the evaluation	Outputs, efficiency, results, effectiveness, impact	*What*
Means of the evaluation	(1. Qualitative, quantitative, mixed methods; 2 data)	*How*
Internal conditional factors	Actors, processes, time, results	*How, when and where*
External conditional factors	Time, context, other contextual factors, potential heterogeneity factors, criteria, rules, standards, understandings, incentives, actions, consequences	*How, when and where*

the available software tools is not assessed. *Daraio* et al. [24.216] point to the need to develop a repository and standards for checking the quality of the available software for running productivity and efficiency analysis.

24.7.3 A Preliminary Checklist

In this chapter we maintain that *standards* and *harmonization* should be applied as well to the methodological dimension of the research assessment process. We propose here to *standardize the methodological steps* according to our framework described in Sect. 24.3.

The preliminary checklist we propose aims to guide the development of econometric models of research assessment, and is reported in Fig. 24.4. It contains questions developers of econometric models should ask themselves and provide answers to, and it makes important analytical distinctions that should be taken into account during the development process.

Table 24.8 lists the main components of the doubly conditional performance evaluation model described in Sect. 24.3. The objective of this table is to facilitate the identification and choice of the main elements and relevant components of the econometric model for research assessment.

Once the econometric model has been developed (according to Fig. 24.4 and Table 24.8) the *sensitivity auditing* checklist of *Saltelli* et al. [24.162], which we report below, should be used to check the robustness of the model and assess the following aspects:

1. Use models to clarify, not to obscure: models as useful tools to represent and clarify reality
2. Adopt an assumption-hunting attitude: listing the underlying assumptions of each approach
3. Detect pseudoscience (uncertainty, spurious decisions, garbage-in garbage-out): make approximations by taking into account data representativeness and role of variables
4. Find sensitive assumptions before they find you: find the critical points in the theoretical framework that deserve attention
5. Aim for transparency: increasing the diffusion of the used model's basic ideas and avoiding jargon
6. Don't do the sums right but do the right sums: concentrate the analysis on the most important components/aspects
7. Focus the analysis: check sensitivity analysis not on one factor at a time but changing the different parameters together.

A second, complementary checklist useful to consider in this context is the *checklist of decision quality control* developed by *Kahneman* et al. [24.217], which is made of the following items:

- Self-interested biases
- Affect heuristic
- Group-think
- Saliency bias (analogy to a memorable case)
- Confirmatory bias
- Availability bias
- Anchoring bias
- Halo effect
- Sunk-cost fallacy, endowment effect
- Overconfidence, planning fallacy
- Optimistic biases, competition neglect
- Disaster neglect
- Loss aversion.

Of course, this is a first attempt and the proposed checklist should be corroborated and tested in the development of practical econometric models. Only by applying and developing/improving the proposed checklist we will be able to reach an acceptable standardized methodology. The collection, elaboration and diffusion of the information requested in the checklist could be very informative for our understanding and a correct development and application of econometrics models.

24.8 Conclusions

In this chapter, we have analyzed what is meant by measuring the productivity of research today, with the advent of the big data era and the fourth (information) revolution [24.5].

We have seen that we need to address the issue of measuring productivity in a broader perspective. We have presented a general framework and a doubly conditional performance assessment model (doubly conditional to the information considered and to those not included in the analysis model).

We have proposed a first checklist for the development of econometric models for evaluation of research.

Our proposed framework has three enabling conditions that foster the connection of the modeling activity with the empirical and policy worlds. They are:

1. *Convergence* (as an evolution of the transdisciplinary approach, which allows for overcoming the traditional paradigms, increasing the dimensional space of thinking)
2. *Mixed methods* (as an intelligent combination of quantitative and qualitative approaches)
3. *Knowledge infrastructures* (as networks of people that interact with artifacts, tools and data infrastructures).

Recent trends in research, economics and econometric theory show the growing need to adopt *broader perspectives*. Such prospects should embrace both the need to develop qualitative-quantitative approaches and to head towards *interdisciplinarity* and *convergence*. Qualitative-quantitative analysis offers strengths that offset the weaknesses of both quantitative and qualitative research [24.218]. Quantitative analysis is weak in understanding the context; qualitative analysis is weak because of personal interpretation difficulty in generalizing. Qualitative-quantitative analysis may act as *a bridge across adversarial divide, between quantitative and qualitative*, may encourages the use of multiple paradigms (beliefs and values), and be *practical* to solve problems, combining inductive and deductive thinking. It allows for the formalization of concepts and measurements. It offers the flexibility of qualitative research. It allows for accountability, and intended and unintended consequences.

Interdisciplinarity leads to the creation of a theoretical, conceptual and methodological identity, hence more coherent and integrated results are obtained. *Convergence* is [24.219]:

> The coming together of insights and approaches from originally distinct fields (...) provides power to think beyond usual paradigms and to approach issues informed by many perspectives instead of few.

Acknowledgments. We warmly thank Léopold Simar and an anonymous reviewer who helped us to greatly improve the content of this chapter. Michele Pezzoni provided helpful suggestions about the literature on determinants of scientific productivity. The financial support of the Italian Ministry of Education and Research (through the PRIN Project N. 2015RJARX7) and of Sapienza University of Rome (through the Sapienza Awards no. PH11715C8239C105) is gratefully acknowledged.

References

24.1 A. Bonaccorsi, C. Daraio: Econometric approaches to the analysis of productivity of R&D systems, production functions and production frontiers. In: *Handbook of Quantitative Science and Technology Research*, ed. by H.F. Moed, W. Glänzel, U. Schmoch (Springer, Dordrecht 2004) pp. 51–74

24.2 C. Daraio: A framework for the assessment of research and its impacts, J. Data Inf. Sci. **2**(4), 7–42 (2017)

24.3 C. Daraio: Assessing research and its impacts: The generalized implementation problem and a doubly-conditional performance evaluation model. In: *Proc. 16th Int. Conf. Scientometr. ISSI* (2017) pp. 1546–1557

24.4 C. Daraio: *The Democratization of Evaluation and Altmetrics*, Technical Report DIAG (Sapienza University of Rome, Rome 2018)

24.5 L. Floridi: *The Fourth Revolution: How the Infosphere is Reshaping Human Reality* (OUP, Oxford 2014)

24.6 W. Glänzel, B. Thijs, K. Debackere: Productivity, performance, efficiency, impact—What do we measure anyway?, Some comments on the paper "A farewell to the MNCS and like size-independent indicators" by Abramo and D'Angelo, J. Informetr. **10**(2), 658–660 (2016)

24.7 L. Waltman: Special section on size-independent indicators in citation analysis, J. Informetr. **10**(2), 645 (2016)

24.8 G. Abramo, C.A. D'Angelo: A farewell to the MNCS and like size-independent indicators, J. Informetr. **10**(2), 646–651 (2016)

24.9 M. Zitt: Paving the way or pushing at open doors? A comment on Abramo and D'Angelo "Farewell to

size-independent indicators", J. Informetr. **2**(10), 675–678 (2016)
24.10 H.F. Moed: *Applied Evaluative Informetrics* (Springer, Cham 2017)
24.11 OECD: *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development*, *The Measurement of Scientific, Technological and Innovation Activities* (OECD, Paris 2015)
24.12 OECD: *Making Open Science a Reality*, OECD Science, Technology and Industry Policy Papers, Vol. 25 (OECD, Paris 2015)
24.13 C. Daraio, L. Simar: *Advanced Robust and Non-Parametric Methods in Efficiency Analysis: Methodology and Applications* (Springer, New York 2007)
24.14 S.K. Mishra: Brief history of production functions, IUP J. Manag. Econ. **8**(4), 6–34 (2010)
24.15 W. Härdle: *Applied Nonparametric Regression* (Cambridge Univ. Press, Cambridge 1992)
24.16 D.J. Henderson, C.F. Parmeter: *Applied Nonparametric Econometrics* (Cambridge Univ. Press, Cambridge 2015)
24.17 R. Färe, S. Grosskopf, C.K. Lovell: *Production Frontiers* (Cambridge Univ. Press, Cambridge 1994)
24.18 H.O. Fried, C.K. Lovell, S.S. Schmidt (Eds.): *The Measurement of Productive Efficiency and Productivity Growth* (Oxford Univ. Press, Oxford 2008)
24.19 H. Varian: *Microeconomic Theory* (W.W. Norton, New York 1992)
24.20 T.C. Koopmans: Efficient allocation of resources, Econometrica **19**(4), 455–465 (1951)
24.21 G. Debreu: The coefficient of resource utilization, Econometrica **19**(3), 273–292 (1951)
24.22 R.W. Shephard: *Theory of Cost and Production Functions* (Princeton Univ. Press, Princeton 1970)
24.23 M.J. Farrell: The measurement of productive efficiency, J. Royal Stat. Soc. Ser. A **120**(3), 253–290 (1957)
24.24 A. Charnes, W.W. Cooper, E. Rhodes: Measuring the efficiency of decision making units, Eur. J. Oper. Res. **2**(6), 429–444 (1978)
24.25 S.C. Kumbhakar, C.A.K. Lovell: *Stochastic Frontier Analysis* (Cambridge Univ. Press, Cambridge 2000)
24.26 P.E. Stephan: The economics of science, J. Econ. Lit. **34**(3), 1199–1235 (1996)
24.27 F. Lissoni, J. Mairesse, F. Montobbio, M. Pezzoni: Scientific productivity and academic promotion: A study on French and Italian physicists, Ind. Corp. Change **20**, 253–294 (2011)
24.28 P.E. Stephan, G. Levin: The critical importance of careers in collaborative scientific research, Rev. Econ. Ind. **79**(1), 45–61 (1996)
24.29 S. Stack: Gender, children and research productivity, Res. High. Educ. **45**, 891–920 (2004)
24.30 Y. Xie, K.A. Shauman: Sex differences in research productivity: New evidence about an old puzzle, Am. Sociol. Rev. **63**(6), 847–870 (1998)
24.31 J. Mairesse, M. Pezzoni: Does gender affect scientific productivity?, Rev. Econ. **66**, 65–113 (2015)
24.32 M. Pezzoni, J. Mairesse, P. Stephan, J. Lane: Gender and the publication output of graduate students: A case study, PLoS ONE **11**, e0145146 (2016)
24.33 B. Bozeman, S. Lee: The Impact of Research Collaboration on Scientific Productivity, Soc. Stud. Sci. **35**(5), 673–702 (2003)
24.34 P. Azoulay, J.S.G. Zivin, J. Wang: *Superstar Extinction*, Working Paper No. 14577 (National Bureau of Economic Research, Cambridge 2008)
24.35 B.A. Jacob, L. Lefgren: The impact of research grant funding on scientific productivity, J. Public Econ. **95**, 1168–1177 (2011)
24.36 L. Turner, J. Mairesse: *Individual Productivity Differences in Public Research: How Important are Non-Individual Determinants? An Econometric Study of French Physicists Publications and Citations* (Centre National de la Recherche Scientifique, Paris 2005) pp. 1986–1997
24.37 M.F. Fox: Publication productivity among scientists: A critical review, Soc. Stud. Sci. **13**, 285–305 (1983)
24.38 J.A.D. Holbrook: Basic indicators of scientific and technological performance, Sci. Public Policy **19**(5), 267–273 (1992)
24.39 R. Johnston: Effects of resource concentration on research performance, High. Educ. **28**, 25–37 (1993)
24.40 P. Ramsden: Describing and explaining research productivity, High. Educ. **28**, 207–226 (1994)
24.41 F. Narin, A. Breitzman: Inventive productivity, Res. Policy **24**(4), 507–519 (1995)
24.42 P.E. Stephan: *How Economics Shapes Science* (Harvard Univ. Press, Cambridge 2012)
24.43 P. Bazeley: Conceptualising research performance, Stud. High. Educ. **35**(8), 889–903 (2010)
24.44 G.S. Åkerlind: An academic perspective on research and being a researcher: An integration of the literature, Stud. High. Educ. **33**(1), 17–31 (2008)
24.45 G. Abramo, C.A. D'Angelo: How do you define and measure research productivity?, Scientometrics **101**(2), 1129–1144 (2014)
24.46 P. Wouters, W. Glänzel, J. Gläser, I. Rafols: The dilemmas of performance indicators of individual researchers—an urgent debate in bibliometrics, ISSI Newsletter **9**(3), 48–53 (2013)
24.47 G. Ruocco, C. Daraio, V. Folli, M. Leonetti: Bibliometric indicators: The origin of their log-normal distribution and why they are not a reliable proxy for an individual scholar's talent, Palgrave Commun. **3**, 17064 (2017)
24.48 A.F. Van Raan: Measuring science. In: *Handbook of Quantitative Science and Technology Research*, ed. by H.F. Moed, W. Glänzel, U. Schmoch (Springer, Dordrecht 2004) pp. 19–50
24.49 M. Gibbons, C. Limoges, H. Nowotny, S. Schwartzman, P. Scott, M. Trow: *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies* (SAGE, London 1994)
24.50 H. Nowotny, P. Scott, M. Gibbons: *Re-Thinking Science: Knowledge and the Public in an Age of Uncertainty* (Polity, Cambridge 2001)

24.51 M. Bucchi: *Beyond Technocracy: Science, Politics and Citizens* (Springer, New York 2009)
24.52 A. Benessia, S. Funtowicz, M. Giampietro, A.G. Pereira, J. Ravetz, A. Saltelli, J.P. van der Sluijs: *Science on the Verge*, The Rightful Place of Science (Consortium for Science, Policy & Outcomes, Tempé 2016)
24.53 M. Bucchi, B. Trench (Eds.): *Routledge Handbook of Public Communication of Science and Technology* (Routledge, Hershley 2014)
24.54 R. Whitley, J. Gläser (Eds.): *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems*, Sociology of the Sciences Yearbook, Vol. 26 (Springer, Dordrecht 2007)
24.55 P. Dahler-Larsen: *The Evaluation Society* (Stanford Univ. Press, Stanford 2011)
24.56 D. Hicks: Evolving regimes of multi-university research evaluation, High. Educ. **57**(4), 393–404 (2009)
24.57 K. Jonkers, T. Zacharewicz: Research performance based funding systems: A comparative assessment. Institute for Prospective Technological Studies, Joint Research Centre. EUR 27837 EN, https://doi.org/10.2791/659483. Available at: http://publications.jrc.ec.europa.eu/repository/bitstream/JRC101043/kj1a27837enn.pdf (last accessed 10 December 2017) (2016)
24.58 C. Daraio, A. Bonaccorsi: Beyond university rankings? Generating new indicators on universities by linking data in open platforms, J. Assoc. Inf. Sci. Technol. **68**(2), 508–529 (2017)
24.59 E. Hazelkorn: Measuring world-class excellence and the global obsession with rankings. In: *A Handbook on Globalization and Higher Education*, ed. by S. Marginson, R. Naidoo, R. King (Edward Elgar, Cheltenham 2011) pp. 497–515
24.60 J. Priem, D. Taraborelli, P. Groth, C. Neylon: Altmetrics: A manifesto, 26 October 2010. http://altmetrics.org/manifesto (last accessed 10 January 2019) (2010)
24.61 J. Priem, P. Groth, D. Taraborelli: The altmetrics collection, PLoS ONE **7**(11), e48753 (2012)
24.62 M. Hilbert, P. López: The world's technological capacity to store, communicate, and compute information, Science **332**(6025), 60–65 (2011)
24.63 S. Pinfield, J. Salter, P.A. Bath, B. Hubbard, P. Millington, J.H. Anders, A. Hussain: Open-access repositories worldwide, 2005–2012: Past growth, current characteristics, and future possibilities, J. Assoc. Inf. Sci. Technol. **65**(12), 2404–2421 (2014)
24.64 J.S.A.H. Katz, D. Hicks: Desktop scientometrics, Scientometrics **38**(1), 141–153 (1997)
24.65 B. Cronin, C.R. Sugimoto: *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact* (MIT Press, Cambridge 2014)
24.66 B. Cronin, C.R. Sugimoto (Eds.): *Scholarly Metrics Under the Microscope: From Citation Analysis to Academic Auditing* (Information Today, Medford 2015)
24.67 M. Bladek: DORA: San Francisco declaration on research assessment, Coll. Res. Libr. News **75**(4), 191–196 (2014)
24.68 D. Hicks, P. Wouters, L. Waltman, S.D. Rijcke, I. Rafols: Bibliometrics: The Leiden Manifesto for research metrics, Nature **520**, 429–431 (2015)
24.69 J. Wilsdon, L. Allen, E. Belfiore, P. Campbell, S. Curry, S. Hill, R. Jones, R. Kain, S. Kerridge, M. Thelwall, J. Tinkler, I. Viney, P. Wouters, J. Hill, B. Johnson: The metric tide: Report of the independent review of the role of metrics in research assessment and management. Available at: http://epc.ac.uk/wp-content/uploads/2015/07/2015-07_metric_tide.pdf (last accessed 10 January 2019) (2015)
24.70 M. Zitt: Les Dérives de l'Évaluation de la Recherche: du bon usage de la bibliométrie (The excesses of research evaluation: The proper use of bibliometrics) by Yves Gingras. Paris: Raisons d'Agir Editions, 2014, J. Assoc. Inf. Sci. Technol. **66**(10), 2171–2176 (2015)
24.71 Y. Gingras: *Bibliometrics and Research Evaluation: Uses and Abuses* (MIT Press, Cambridge 2016)
24.72 R. Benedictus, F. Miedema, M.W. Ferguson: Fewer numbers, better science, Nature **538**(7626), 453–455 (2016)
24.73 P. Stephan, R. Veugelers, J. Wang: Blinkered by bibliometrics, Nature **544**(7651), 411–412 (2017)
24.74 H.F. Moed, G. Halevi: Multidimensional assessment of scholarly research impact, J. Assoc. Inf. Sci. Technol. **66**(10), 1988–2002 (2015)
24.75 W. Glänzel: The need for standards in bibliometric research and technology, Scientometrics **35**(2), 167–176 (1996)
24.76 C. Daraio, W. Glänzel: Grand challenges in data integration–state of the art and future perspectives: An introduction, Scientometrics **108**(1), 391–400 (2016)
24.77 W. Glänzel, H. Willems: Towards standardisation, harmonisation and integration of data from heterogeneous sources for funding and evaluation purposes, Scientometrics **106**(2), 821–823 (2016)
24.78 M. Nielsen: *Reinventing Discovery: The New Era of Networked Science* (Princeton Univ. Press, Princeton 2012)
24.79 C.L. Borgman: *Big data, little data, no data: Scholarship in the networked world* (MIT Press, Cambridge 2015)
24.80 J. Stewart: Models of priority-setting for public sector research, Res. Policy **24**(1), 115–126 (1995)
24.81 R. Moscati: Transforming a centralised system of higher education: Reform and academic resistance in Italy. In: *From Governance to Identity, A Festschrift for Mary Henkel*, ed. by A. Amaral, I. Bleiklie, C. Musselin (Springer, Dordrecht 2008)
24.82 S. Boffo, P. Dubois, R. Moscati: Changes in university governance in France and in Italy, Tert. Educ. Manag. **14**(1), 13–26 (2008)
24.83 G. Rebora, M. Turri: The UK and Italian research assessment exercises face to face, Res. Policy **42**(9), 1657–1666 (2013)

24.84 C. Daraio: Non-parametric methods and higher education. In: *Encyclopedia of International Higher Education Systems and Institutions*, ed. by P. Teixeira, J. Shin (Springer, Dordrecht 2018)
24.85 R. Frisch: Editor's Note, Econometrica **1**(1), 1–4 (1936)
24.86 P.A. Samuelson, T.C. Koopmans, J.R.N. Stone: Report of the evaluative committee for Econometrica, Econometrica **22**(2), 141–146 (1954)
24.87 J. Geweke, J. Horowitz, M.H. Pesaran: Econometrics. In: *The New Palgrave Dictionary of Economics*, 2nd edn., ed. by S. Durlauf, L. Blume, U. Schmoch (Palgrave Macmillan, Basingstoke 2008)
24.88 W.J. Baumol, A.S. Blinder: *Economics: Principles and Policy*, 12th edn. (South-Western Cengage Learning, Boston 2012)
24.89 J.D. Gibbons, S. Chakraborti: Nonparametric statistical inference. In: *International Encyclopedia of Statistical Science*, ed. by M. Lovric (Springer, Berlin 2011) pp. 977–979
24.90 B. Efron, R.J. Tibshirani: *An Introduction to the Bootstrap* (CRC, Boca Raton 1994)
24.91 D.P. Kroese, T. Brereton, T. Taimre, Z.I. Botev: Why the Monte Carlo method is so important today, Wiley Interdiscip. Rev. Comput. Stat. **6**(6), 386–392 (2014)
24.92 P.F. Drucker, J.A. Maciariello: *Management* (Collins, New York 2008), Revised edition
24.93 M.E. Porter: *Competitive Advantage: Creating and Sustaining Superior Performance* (Free Press, New York 1985)
24.94 M.E. Porter: *Competitive Advantage of Nations: Creating and Sustaining Superior Performance* (Simon and Schuster, New York 2011)
24.95 M.E. Porter, V.E. Mirrar: How information gives you competitive advantage, Harv. Bus. Rev. **64**(4), 149–160 (1985)
24.96 T.H. Davenport: Competing on analytics, Harv. Bus. Rev. **84**(1), 98 (2006)
24.97 T. Zenger: *Beyond Competitive Advantage* (Harvard Business Review, Boston 2016)
24.98 D. Besanko, D. Dranove, M. Shanley, S. Schaefer: *Economics of Strategy* (Wiley, New Jersey 2009)
24.99 B.M. Frischmann: *Infrastructure: The Social Value of Shared Resources* (Oxford Univ. Press, Oxford 2012)
24.100 A. Tversky, D. Kahneman: Judgment under uncertainty: Heuristics and biases, Science **185**(4157), 1124–1131 (1974)
24.101 H.A. Simon: *The Sciences of the Artificial* (MIT Press, Cambridge 1969)
24.102 H.A. Simon: *Models of Bounded Rationality* (MIT Press, Cambridge 1982)
24.103 R. Simons, A. Dávila, R.S. Kaplan: *Performance Measurement & Control Systems for Implementing Strategy* (Prentice Hall, Upper Saddle River 2000) p. 203
24.104 D. Kahneman, P. Egan: *Thinking, Fast and Slow* (Farrar, Straus and Giroux, New York 2011)
24.105 J. Friedman, T. Hastie, R. Tibshirani: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, New York 2009)
24.106 M. Mezard, A. Montanari: *Information, Physics, and Computation* (Oxford Univ. Press, Oxford 2009)
24.107 K.P. Murphy: *Machine Learning: A Probabilistic Perspective* (MIT Press, Cambridge 2012)
24.108 D. Barber: *Bayesian Reasoning and Machine Learning* (Cambridge Univ. Press, Cambridge 2012)
24.109 M.I. Jordan, T.M. Mitchell: Machine learning: Trends, perspectives, and prospects, Science **349**(6245), 255–260 (2015)
24.110 A. Graves, G. Wayne, M. Reynolds, T. Harley, I. Danihelka, A. Grabska-Barwinska, A.P. Badia: Hybrid computing using a neural network with dynamic external memory, Nature **538**(7626), 471–476 (2016)
24.111 R. Layard, S. Glaister: *Cost–Benefit Analysis* (Cambridge Univ. Press, Cambridge 1994)
24.112 A.E. Boardman, D.H. Greenberg, A.R. Vining, D.L. Weimer: *Cost–Benefit Analysis: Concepts and Practice*, 4th edn. (Prentice Hall, Upper Saddle River 2011)
24.113 D. Weimer (Ed.): *Cost–Benefit Analysis and Public Policy* (Wiley, Blackwell 2008)
24.114 A.M. Ghalayini, J.S. Noble: The changing basis of performance measurement, Int. J. Oper. Prod. Manag. **16**(8), 63–80 (1996)
24.115 R.G. Eccles: The performance measurement manifesto, Harv. Bus. Rev. **69**(1), 131–137 (1991)
24.116 A. Neely (Ed.): *Business Performance Measurement: Theory and Practice* (Cambridge Univ. Press, Cambridge 2002)
24.117 R.S. Kaplan, D.P. Norton: The Balanced Scorecard—Measures that Drive Performance, Harv. Bus. Rev. **70**, 71–79 (1992)
24.118 A.A. Atkinson, J.H. Waterhouse, R.B. Wells: A stakeholder approach to strategic performance measurement, MIT Sloan Manag. Rev. **38**(3), 25 (1997)
24.119 A.D. Neely, C. Adams, M. Kennerley: *The Performance Prism: The Scorecard for Measuring and Managing Business Success* (Financial Times/Prentice Hall, London 2002)
24.120 E. Ferlie, L. Fitzgerald, A. Pettigrew: *The New Public Management in Action* (Oxford Univ. Press, Oxford 1996)
24.121 J.E. Lane: *New Public Management: An Introduction* (Routledge, London 2002)
24.122 R.S. Kaplan: Strategic performance measurement and management in nonprofit organizations, Nonprofit Manag. Leadersh. **11**(3), 353–370 (2001)
24.123 W. Van Dooren, G. Bouckaert, J. Halligan: *Performance Management in the Public Sector* (Routledge, London 2015)
24.124 R.D. Behn: The new public management paradigm and the search for democratic accountability, Int. Public Manag. J. **1**(2), 131–164 (1998)

24.125 R.C. Box, G.S. Marshall, B.J. Reed, C.M. Reed: New public management and substantive democracy, Public Adm. Rev. **61**(5), 608–619 (2001)
24.126 P. Dunleavy, H. Margetts, S. Bastow, J. Tinkler: New public management is dead—long live digital-era governance, J. Public Adm. Res. Theory **16**(3), 467–494 (2006)
24.127 J. O'Flynn: From new public management to public value: Paradigmatic change and managerial implications, Aust. J. Public Adm. **16**(3), 353–366 (2007)
24.128 S. Van Thiel, F.L. Leeuw: The performance paradox in the public sector, Public Perform. Manag. Rev **25**(3), 267–281 (2002)
24.129 L. Bryl: Intellectual capital measurement. In: *Encyclopedia of Information Science and Technology*, 4th edn., ed. by M. Khosrow-Pour (IGI Global, Hershley 2018) pp. 5056–5066
24.130 A. Tiwana: *The Knowledge Management Toolkit: Practical Techniques for Building a Knowledge Management System* (Prentice Hall PTR, Upper Saddle River 2000)
24.131 M. Alavi, D.E. Leidner: Knowledge management and knowledge management systems: Conceptual foundations and research issues, MIS Quarterly **25**(1), 107–136 (2001)
24.132 J. Liebowitz (Ed.): *Knowledge Management Handbook* (CRC, Boca Raton 2012)
24.133 A. Bounfour, L. Edvinsson (Eds.): *Intellectual Capital for Communities* (Routledge, Abingdon-on-Thames 2012)
24.134 J. Guthrie, J. Dumay: New frontiers in the use of intellectual capital in the public sector, J. Intellect. Cap. **16**(2), 258–266 (2015)
24.135 J. Dumay, J. Guthrie, P. Puntillo: IC and public sector: A structured literature review, J. Intellect. Cap. **16**(2), 267–284 (2015)
24.136 G. Secundo, R. Lombardi, J. Dumay: Intellectual capital in education, J. Intellect. Cap. **19**(1), 2–9 (2018)
24.137 A. Pulic: *The Principles of Intellectual Capital Efficiency—A Brief Description* (Croatian Intellectual Capital Center, Zagreb Center 2008)
24.138 G. Iazzolino, D. Laise: Value added intellectual coefficient (VAIC): A methodological and critical review, J. Intellect. Cap. **14**(4), 547–563 (2013)
24.139 D. Helbing, S. Bishop, R. Conte, P. Lukowicz, J.B. McCarthy: FuturICT: Participatory computing to understand and manage our complex world in a more sustainable and resilient way, Eur. Phys. J. Spec. Top. **214**(1), 11–39 (2012)
24.140 J.A. Doornik, D.F. Hendry: Statistical model selection with "Big Data", Cogent Econ. Finance **3**(1), 1045216 (2015)
24.141 H. Ekbia, M. Mattioli, I. Kouper, G. Arave, A. Ghazinejad, T. Bowman, C.R. Sugimoto: Big data, bigger dilemmas: A critical review, J. Assoc. Inf. Sci. Technol. **66**(8), 1523–1545 (2015)
24.142 F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, D.R. White: Economic networks: What do we know and what do we need to know?, Adv. Complex Syst. **12**(04n05), 407–422 (2009)
24.143 F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, A. Vespignani, D.R. White: Economic networks: The new challenges, Science **325**(5939), 422–425 (2009)
24.144 W. Elsner, T. Heinrich, H. Schwardt: *The Microeconomics of Complex Economies: Evolutionary, Institutional, Neoclassical, and Complexity Perspectives* (Academic Press, Oxford 2015)
24.145 A. Kirman: Networks: A paradigm shift for economics? In: *The Oxford Handbook of the Economics of Networks*, ed. by Y. Bramoullé, A. Galeotti, B.W. Rogers (Oxford Univ. Press, Oxford 2016)
24.146 M.P. Mandell, R. Keast: Evaluating the effectiveness of interorganizational relations through networks: Developing a framework for revised performance measures, Public Manag. Rev. **10**(6), 715–731 (2008)
24.147 A. Golan: Information and entropy econometrics, A review and synthesis, Found. Trends Econ. **2**(12), 1–145 (2008)
24.148 G.G. Judge, R.C. Mittelhammer: *An Information Theoretic Approach to Econometrics* (Cambridge Univ. Press, Cambridge 2011)
24.149 N. Georgescu-Roegen: The economics of production, Am. Econ. Rev. **60**(2), 1–9 (1970)
24.150 N. Georgescu-Roegen: *Entropy Law and the Economic Process* (Cambridge Univ. Press, Cambridge 1971)
24.151 N. Georgescu-Roegen: Process analysis and the neoclassical theory of production, Am. J. Agric. Econ. **54**(2), 279–294 (1972)
24.152 N. Georgescu-Roegen: Methods in economic science, J. Econ. Issues **13**(2), 317–328 (1979)
24.153 M. Morroni: *Production Process and Technical Change* (Cambridge Univ. Press, Cambridge 1992)
24.154 M. Morroni: Production of commodities by means of processes. The flow-fund model, input-output relations and the cognitive aspects of production, Struct. Change Econ. Dyn. **29**, 5–18 (2014)
24.155 M. Morroni: *Knowledge, Scale and Transactions in the Theory of the Firm* (Cambridge Univ. Press, Cambridge 2006) pp. 1–358
24.156 G. Fioretti: The production function, Physica A **374**(2), 707–714 (2007)
24.157 C. Kao: *Network Data Envelopment Analysis, Foundations and Extensions* (Springer, Cham 2017)
24.158 C. Daraio, R. Färe, S. Grosskopf, M.G. Izzo, L. Leuzzi, G. Ruocco, M. Bostian: Inference for nonparametric productivity networks: A pseudo-likelihood approach, paper presented at the CM-Stat Conference, London (2017)
24.159 B. Bozeman, J. Melkers (Eds.): *Evaluating R&D Impacts: Methods and Practice* (Springer, New York 1993)
24.160 S.B. Khandker, G. Koolwal, H. Samad: *Handbook on Impact Evaluation: Quantitative Methods and Practices* (The World Bank, Washington DC 2010)
24.161 A. Saltelli, A. Guimaraes-Pereira, J.P. van der Sluijs, S. Funtowicz: What do I make of your latinorum? Sensitivity auditing of mathematical modelling, Int. J. Foresight Innov. Policy **9**, 2–3 (2013)

24.162 A. Saltelli, S. Funtowicz: When all models are wrong, Issues Sci. Technol. **30**(2), 79–85 (2014)

24.163 A. Saltelli, S. Funtowicz: Evidence-based policy at the end of the Cartesian dream: The case of mathematical modelling. In: *Science, Philosophy and Sustainability: The End of the Cartesian Dream*, Routledge Explorations in Sustainability and Governance, ed. by Â.G. Pereira, S. Funtowicz (Routledge, New York 2015) pp. 147–162

24.164 Z. Griliches, J. Mairesse: Production functions: The search for identification. In: *Practising Econometrics: Essays in Method and Application*, ed. by Z. Griliches (Edward Elgar, Cheltenham 1998) pp. 383–411

24.165 A.R. Hall: *Generalized Method of Moments* (Oxford Univ. Press, Oxford 2005)

24.166 W.H. Greene: *Econometric Analysis* (Pearson Education, Upper Saddle River 2003)

24.167 A. Pagan, A. Ullah: *Nonparametric Econometrics* (Cambridge Univ. Press, Cambridge 1999)

24.168 D. Aigner, C.K. Lovell, P. Schmidt: Formulation and estimation of stochastic frontier production function models, J. Econometr. **6**(1), 21–37 (1977)

24.169 A. Kneip, L. Simar: A general framework for frontier estimation with panel data, J. Prod. Anal. **7**(2/3), 187–212 (1996)

24.170 P. Hall, L. Simar: Estimating a changepoint, boundary or frontier in the presence of observation error, J. Am. Stat. Assoc. **97**, 523–534 (2002)

24.171 L. Simar: How to improve the performances of DEA/FDH estimators in the presence of noise, J. Product. Anal. **28**, 183–201 (2007)

24.172 S.C. Kumbhakar, B.U. Park, L. Simar, E.G. Tsionas: Nonparametric stochastic frontiers: A local likelihood approach, J. Econ. **137**(1), 1–27 (2007)

24.173 L. Simar, V. Zelenyuk: Stochastic FDH/DEA estimators for frontier analysis, J. Product. Anal. **36**, 1–20 (2011)

24.174 B. Park, L. Simar, V. Zelenyuk: Categorical data in local maximum likelihood: Theory and application to productivity analysis, J. Product. Anal. **43**, 199–214 (2015)

24.175 J.P. Florens, L. Simar, I. Van Keilegom: *Estimation of the Boundary of a Variable Observed with Symmetric Error*, Discussion paper 2018/08, ISBA, UCL (Université Catholique de Louvain, Louvain-la-Neuve 2018)

24.176 W. Härdle: *Applied Nonparametric Regression* (Cambridge Univ. Press, Cambridge 1990)

24.177 J.P. Florens, L. Simar: Parametric approximations of non-parametric frontiers, J. Econ. **124**(1), 91–116 (2005)

24.178 L. Simar, P.W. Wilson: Estimation and inference in nonparametric frontier models: Recent developments and perspectives, Found. Trends Econ. **5**(3/4), 183–337 (2013)

24.179 L. Simar, P.W. Wilson: Statistical approaches for non-parametric frontier models: A guided tour, Int. Stat. Rev. **83**(1), 77–110 (2015)

24.180 C. Mastromarco, L. Simar, P.W. Wilson: Nonparametric statistical analysis of production. In: *Handbook of Economic Performance Analysis*, ed. by W. Greene, T. ten Raa (Palgrave Macmillan, Basingstoke 2018)

24.181 Z. Griliches: Economic Data Issues. In: *Handbook of Econometrics*, Vol. 3, ed. by Z. Griliches, M. Intriligator (North Holland, Amsterdam 1986) pp. 1465–1514

24.182 Z. Griliches: Productivity, R&D, and the data constraint, Am. Econ. Rev. **84**(1), 1–23 (1994)

24.183 C. Parent, S. Spaccapietra: Database integration: The key to data interoperability. In: *Advances in Object-Oriented Data Modeling*, ed. by M.P. Papazoglou, Z. Zari (MIT Press, Cambridge 2000) pp. 221–253

24.184 M. Luwel: The use of input data in the performance analysis of R&D systems. In: *Handbook of Quantitative Science and Technology Research*, ed. by H.F. Moed, W. Glänzel, H. Schmoch (Springer, Dordrecht 2004) pp. 315–338

24.185 D. Aksnes, G. Sivertsen, T.N. van Leeuwen, K.K. Wendt: Measuring the productivity of national R&D systems: Challenges in cross-national comparisons of R&D input and publication output indicators, Sci. Public Policy **44**(2), 246–258 (2017)

24.186 A. Bonaccorsi, C. Daraio: *Universities and Strategic Knowledge Creation: Specialization and Performance in Europe* (Edward Elgar, Cheltenham 2007)

24.187 A. Bonaccorsi, C. Daraio, B. Lepori, S. Slipersaeter: Indicators on individual higher education institutions: addressing data problems and comparability issues, Res. Eval. **16**(2), 66–78 (2007)

24.188 C. Daraio, A. Bonaccorsi, A. Geuna, B. Lepori, L. Bach, P. Bogetoft, M.F. Cardoso, E. Castro-Martinez, G. Crespi, I. Fernandez de Lucio, H. Fried, A. Garcia-Aracil, A. Inzelt, B. Jongbloed, G. Kempkes, P. Llerena, M. Matt, M. Olivares, C. Pohl, T. Raty, M.J. Rosa, C.S. Sarrico, L. Simar, S. Slipersaeter, P.N. Teixeira, P. Vanden Eeckaut: The European university landscape: A micro characterization based on evidence from the Aquameth project, Res. Policy **40**(1), 148–164 (2011)

24.189 C. Daraio, M. Scannapieco, T. Catarci, L. Simar: ETER Data Quality Report. Available at: https://www.eter-project.com/uploads/assets/pdf/ETER_quality_report_2017.pdf (last accessed 11 January 2019) (2017).

24.190 G.E.P. Box: Science and Statistics, J. Am. Stat. Assoc. **71**, 791–799 (1976)

24.191 J.H. Stock: The other transformation in econometric practice: Robust tools for inference, J. Econ. Perspect. **24**(2), 83–94 (2010)

24.192 Q. Li, J.S. Racine: *Nonparametric Econometrics: Theory and Practice* (Princeton Univ. Press, Princeton 2007)

24.193 H.R. Varian: Big data: New tricks for econometrics, J. Econ. Perspect. **28**(2), 3–27 (2014)

24.194 F.J. Alexander, A. Hoisie, A. Szalay: Big data, Comput. Sci. Eng. **13**(6), 10–13 (2011)

24.195 J.B. Cushing: Beyond Big Data?, Comput. Sci. Eng. **15**(5), 4–5 (2013)

24.196 L. Einav, J. Levin: Economics in the age of big data, Science **346**(6210), 1243089 (2014)

24.197 R. Rossi, K. Hirama: Characterizing big data management, Issues Inf. Sci. Inf. Technol. **12**, 165–180 (2015)
24.198 S.F. Wamba, S. Akter, A. Edwards, G. Chopin, D. Gnanzou: How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study, Int. J. Prod. Econ. **165**, 234–246 (2015)
24.199 H.R. Varian: Beyond big data, Bus. Econ. **49**(1), 27–31 (2014)
24.200 C. Daraio: Challenges of big data for economic modeling and management. Tools from efficiency analysis, sensitivity analysis, sensitivity auditing and physics of complex systems. In: *Proc. Workshop 10/11, 2015 DIAG, Sapienza Univ. Rome*, ed. by C. Daraio (Efesto Edizioni, Roma 2016)
24.201 D. Acemoglu, P. Aghion, F. Zilibotti: Vertical integration and distance to frontier, J. Eur. Economic Assoc. **1**(2/3), 630–638 (2003)
24.202 D. Acemoglu, P. Aghion, F. Zilibotti: Distance to frontier, selection, and economic growth, J. Eur. Econ. Assoc. **4**(1), 37–74 (2006)
24.203 J. Vandenbussche, P. Aghion, C. Meghir: Growth, distance to frontier and composition of human capital, J. Econ. Growth **11**(2), 97–127 (2006)
24.204 C.M. Chen, M.A. Delmas, M.B. Lieberman: Production frontier methodologies and efficiency as a performance measure in strategic management research, Strateg. Manag. J. **36**(1), 19–36 (2015)
24.205 J.E. Galán, H. Veiga, M.P. Wiper: Bayesian estimation of inefficiency heterogeneity in stochastic frontier models, J. Product. Anal. **42**(1), 85–101 (2014)
24.206 L. Bădin, C. Daraio, L. Simar: How to measure the impact of environmental factors in a non-parametric production model, Eur. J. Oper. Res. **223**(3), 818–833 (2012)
24.207 L. Bădin, C. Daraio, L. Simar: Explaining inefficiency in non-parametric production models: The state of the art, Ann. Oper. Res. **214**(1), 5–30 (2014)
24.208 C. Daraio, L. Simar: Directional distances and their robust versions: Computational and testing issues, Eur. J. Oper. Res. **237**(1), 358–369 (2014)
24.209 C. Daraio, L. Simar, P.W. Wilson: Central limit theorems for conditional efficiency measures and tests of the "separability" condition in nonparametric, two-stage models of production, Econ. J. **21**(2), 170–191 (2018)
24.210 C.F. Parmeter, A.T. Wan, X. Zhang: A model averaging stochastic frontier estimator. Paper presented at the NAPW 2016 Quebec City, June 2016
24.211 C.J. O'Donnell: Using information about technologies, markets and firm behaviour to decompose a proper productivity index, J. Econ. **190**(2), 328–340 (2016)
24.212 P. Bogetoft: *Performance Benchmarking: Measuring and Managing Performance* (Springer, New York 2013)
24.213 L. Simar, A. Vanhems, I. Van Keilegom: Unobserved heterogeneity and endogeneity in nonparametric frontier estimation, J. Econ. **190**(2), 360–373 (2016)
24.214 C.F. Parmeter, S.C. Kumbhakar: Efficiency analysis: A primer on recent advances, Found. Trends Econ. **7**(3/4), 191–385 (2014)
24.215 R.C. Sickles, V. Zelenyuk: *Measurement of Productivity and Efficiency: Theory and Practice* (Cambridge Univ. Press, Cambridge 2018), forthcoming in 2019
24.216 C. Daraio, K. Kerstens, T. Nepomuceno, R.C. Sickles: Productivity and efficiency analysis software: An exploratory bibliographical survey of the options, J. Econ. Surv. **33**(1), 85–100 (2019)
24.217 D. Kahnemann, D. Lovallo, O. Sibony: Before you make that big decision, Harv. Bus. Rev. **89**(6), 50–60 (2011)
24.218 J.W. Creswell, P.V.L. Clark: *Designing and conducting mixed methods research* (Sage, Thousand Oaks 2011)
24.219 National Research Council: *Convergence: Facilitating Transdisciplinary Integration of Life Sciences, Physical Sciences, Engineering, and Beyond* (National Academies Press, Washington DC 2014)

Cinzia Daraio

Department of Computer, Control, and Management Engineering
Sapienza University of Rome
Rome, Italy
daraio@diag.uniroma1.it

Cinzia Daraio is Associate Professor of Management Engineering at the Sapienza University of Rome, where she teaches productivity and efficiency analysis, quantitative models for economic analysis and management, and economics and business organization.
She specializes in science and technology indicators, higher education microdata, and methodological and empirical studies in productivity and efficiency analysis.
She is a member of the editorial board of *Scientometrics*, *Journal of Productivity Analysis, International Transactions in Operational Research*, and others.

25. Developing Current Research Information Systems (CRIS) as Data Sources for Studies of Research

Gunnar Sivertsen

Current research information systems (CRIS) are increasingly being used to standardize and ease documentation, communication, and administration of research. With broad coverage and sufficient completeness, data quality, and standardization, CRIS systems can also be used as data sources for studies of research. Making CRIS interoperable and comparable across institutions and countries is necessary for the further development of CRIS for research purposes. Integration of CRIS for administrative purposes is already on the European agenda. This chapter focuses on challenges and solutions to the development of internationally integrated CRIS. Most of the remaining challenges are not related to technical solutions, but to an efficient sharing and use of *contents*. The chapter starts with the situation at the international level before it moves on to an example of CRIS at the national level to describe challenges and possible solutions even more concretely. The last section of the chapter provides examples of the type of studies that can be performed if progress is made for internationally integrated CRIS.

25.1 Current Research Information Systems

Current research information systems (CRIS) are databases or other information systems used within and among research organizations to store, manage, and exchange data for documentation, communication, and administration of research activities. CRIS usually contain information about researchers and research groups, their projects, funding, outputs, and outcomes. In the most advanced versions, CRIS help produce integrated data for what used to be documents for separate purposes, such as individual applications for funding, institutional annual reports, project reports, CVs, publications lists, profiles of research groups, project reports, information for media and the general public, etc. Searchable bibliographic references may lead on to full texts in local repositories.

So far, most CRIS operate at the institutional level only. They are, however, becoming nationally integrated in an increasing number of countries, and there are initiatives to integrate them internationally as well. We will focus particularly on integration of CRIS in this chapter because it represents an important condition for using CRIS for more general purposes. With integrated CRIS providing data that are structured and quality assured for statistical purposes, research performing and funding organizations may also use CRIS for monitoring and evaluating research activities and outputs, allocating funding, supporting decision making on their policies and strategies, tracking researchers' careers, and describing their systemic role to policy-makers, stakeholders, and the public. With broad coverage and

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_25

sufficient completeness, data quality and standardization, CRIS systems can also be used as data sources for *studies of research*.

We are, however, not there yet. In addition to introducing CRIS developments and giving examples of their potential for providing data for studies of research, this chapter will focus on challenges and possible solutions related to CRIS integration. We observe the same main problems here as have been observed more generally by *Daraio* and *Glänzel* [25.1] in relation to data integration for research administration, evaluation, and policy. The problems are data quality, comparability, standardization, interoperability, modularization, classifications and concordance tables, comprehensiveness and completeness, and shared schedules for updating. The chapter presents initiatives to solve such problems and summarizes them with a few policy recommendations as part of the conclusions.

With the main problems solved, using CRIS data can overcome several of the limitations presently encountered in data sources for research administration, evaluation, and studies of research. A CRIS is not the same as a repository of publications or a bibliographic database. A CRIS unites all information sources that are relevant for the administration of research activities in one dynamic interrelated system. It is by constantly interrelating bibliographic information and other types of information representing the factors influencing scientific production that a CRIS can break new paths in studies of research activities. A CRIS has information about identifiable persons (not only authors) and institutional affiliations, titles, and positions (not only published addresses), as well as more complete economic information than is available in funding acknowledgements in publications. Such extended information may lead to other data (personal variables, backgrounds, resources, projects, networks, memberships) within or outside the CRIS system, which may serve more general studies of research based on social science methods.

Another strength of CRIS from the same perspective is the possibility of completeness in the coverage of the published literature. Coverage can go beyond the existing bibliographic data sources such as the Web of Science (WoS) and Scopus [25.2]. This may be important in research areas such as engineering science, social sciences, and humanities. While most types of bibliometric analysis, such as studies of collaboration and output profiles, can be performed by using CRIS data alone, there is both a limitation and a new option for citation analysis. The limitation is that CRIS do not index citations themselves. The data will need to be matched to another data source. The new option, however, is that CRIS makes it easier to attribute non-indexed publications and citations provided by Google Scholar or Microsoft Academic to persons and affiliations.

The use of CRIS for local institutional purposes has become widespread during the last decade and is now served by a commercial market of several professional providers of CRIS solutions. The development of CRIS for studies of research, however, is still in an early phase. The lack of data sharing options and comparability is still a major limitation. The data need to be available and comparable across local CRIS. This is so far only the case in a few countries that have managed to establish a national, non-commercial CRIS or a system for aggregating data from local CRIS. We will present examples of how such national CRIS are built and integrated and how they have already been used for studies of research, in Sects. 25.9–25.11 of this chapter.

As explained above, the major advantage of using data from CRIS in administration, evaluation, and studies of research would be realized only after establishing internationally integrated CRIS with comparable data. The feasibility of establishing a European integrated research information infrastructure, mainly for administrative purposes, was demonstrated in a report to the European Parliament [25.3]. In 2016, Science Europe published a *position statement on research information systems* inviting all research organizations to develop resilient information systems by adopting certain core principles and technical recommendations [25.4]. These two documents represent the point of departure for this chapter, and we will start by presenting them in more detail in Sects. 25.2–25.3.

The following sections will focus on what is needed for the *further development* of CRIS as data sources for studies of research. Several initiatives and solutions that may contribute to an internationally integrated CRIS will be described, starting with more general examples in Sects. 25.4–25.8. In Sects. 25.9–25.10, the chapter then moves on to an example of a CRIS at the national level to describe challenges and possible solutions even more concretely. Section 25.11 provides examples of the type of studies that can be performed if progress is made for internationally integrated CRIS, while Sect. 25.12 summarizes the chapter and discusses ways to move forward at the policy level.

The focus of the chapter will be less on *technical* solutions than on challenges and possible solutions related to creating comparable *contents* in CRIS. In our view, solutions that may provide the technical feasibility of an internationally integrated CRIS are already available to a large extent. As examples, an international standard for the technical format of a CRIS, the Common European Research Information Format (CERIF) and an international researcher identifier, the Open Researcher and Contributor ID (ORCID) are already in place. Such technical solutions are well documented elsewhere and will only be mentioned in the following. The emphasis of the chapter will be on challenges and possible solutions for CRIS data integration that are related to defining and standardizing data, ensuring their completeness and quality, and agreeing on sharing them. Here, the solutions will depend on the involvement of expertise in data analysis for research administration, evaluation and policy—and ultimately also on international policy level agreements.

25.2 The Need for Top-Down Coordination

In a report to the European Parliament in 2014 on *Measuring scientific performance for improved policy making* [25.3], the mandate was to explore the feasibility of a transnational system for collecting and monitoring research performance data (on inputs, outputs, and productivity) in Europe. The conclusion was that a European integrated research information infrastructure is not only technically feasible, but needed to achieve several benefits (cited from the document):

- For research institutions: The possibility directly to compare and benchmark research performance with other institutions in Europe, taking into consideration the different missions of the institutions, their research infrastructures and national environments, thus improving the awareness of the institution's positioning in the European research landscape—beyond the analysis of bibliometrics.
- For national funding agencies and policy makers: A comprehensive view of the complementarities of national research strategies versus other countries and the European Commission; improved basis for comparisons and benchmarking of national research performance with other countries, in line with the proper needs.
- For the European Commission: Improved efficiency in the collection of micro-data, improving data availability, reducing duplicates and enhancing the sustainability of data collection efforts.
- For the research performance assessment community at large: The basis for an improved understanding of knowledge exchange mechanisms in the European research system, providing a comprehensive view on input and outputs.

The benefit for studies of research could be added.

The study concluded that a possible lack of comparability of data in different systems does not constitute a major technical problem thanks to the use of semantic tools. This conclusion is perhaps a bit optimistic, although it refers to successful international initiatives and solutions such as CERIF, CASRAI, and ORCID, which we will present in the next section.

The most important recommendation of the study, in our view, is not to substitute for existing national research information systems, but to build an additional layer on top of them, which would comprise a distributed infrastructure, inter-connecting existing national research information systems. After the study was completed, the example of the VIRTA Research Information Management Services Finland [25.5], which we discuss in Sects. 25.6 and 25.7, has in practice demonstrated support for this recommendation. The same VIRTA project could now be used to test another conclusion from the study, namely that the costs of developing the integrating system would be relatively limited. As far as we can see from the VIRTA project, developing technical solutions seems to be less expensive than working for comparability of contents across countries. This learning might become tested once more in the experimentation with CRIS based on CERIF within the OpenAIRE project, see Sect. 25.3.

A third conclusion of the study is that, looking back at experiences so far, the distributed infrastructure for integration of CRIS will not be the result of a bottom-up process. European coordination is needed. Although we completely agree with this conclusion, it should be added that no CRIS would exist without local initiatives and developments. Moreover, internationally integrated CRIS will still need to serve local purposes and save labor–all the way down to the individual researcher. A bottom-up engagement will be needed in the implementation of the international infrastructure and its development towards increased usefulness for multiple purposes.

25.3 Towards Internationally Integrated CRIS

The conclusions of the report presented above are representative of a trend towards making existing or emerging CRIS interoperable rather than creating new databases. Only a decade ago, in 2009, the research councils of four European countries collaborated with the European Research Council on initiating a project that was typically named *Towards a bibliometric database for the social sciences and humanities—A European scoping project* [25.6]. One of the recommendations, to build a new, separate database for the social sciences and humanities, has not been followed up. Another recommendation, which has proved to be more forward-looking, was to build on existing or emerging national databases and standardize their data.

A similar recommendation, based on a more recent European project [25.7], resulted in a *Position statement on research information systems* by *Science Europe* [25.4], the new member organization for research funding organizations in Europe. The new document neither uses the word `database` nor expresses any limitation regarding area of research or level of coverage (local, national, institutional). Reflecting the present technical state-of-the art, with new options for big data exchange and with the agenda of open science, it simply *invites all research organizations to develop resilient research information systems* by following four core principles and some specific recommendations that may enhance their interoperability. The four core principles are (cited from the document):

- *Flexibility*. Research information systems should be flexible enough to allow for extensions in terms of the data objects covered, their definitions, metadata, and use of external data sources.
- *Openness*. Research information systems' data should be available for external use—in line with the principle *as open as possible, as closed as necessary* and EU Directive 2013/37/EU1—and their processing should never require the loss of ownership in underlying raw data by the originating institution.
- *FAIRness*. Research information systems should foster the *f*indability, *a*ccessibility, *i*nteroperability, and *r*eusability of the data that they store by implementing the FAIR guiding principles for research activity data [25.8]. The application of these principles is meant to cover all components of the research process, not only publications and their underlying data, to ensure transparency, reproducibility, and reusability of research.
- *Data entry minimization*. Research information systems should minimize the need for entering data and facilitate the reuse of data entered manually, in line with the motto *enter once, reuse multiple times*.

The statement furthermore recommends making use of available resources and initiatives towards standardization and interoperability. Here, we will only briefly mention five of these recommended resources, since they are already well documented online and in the literature. In our view, these already implemented resources solve different, limited standardization tasks and will be useful as parts of a more general solution:

- *CASRAI*. CASRAI is an international, non-profit organization working for standard information agreements among research information users. CASRAI agreements cover all types of information in the management of research.
- *CERIF* is a data model for building a CRIS with data on projects, people, organizations, publications, patents, products, services, and facilities (equipment in particular) with role-based, temporally-bound relationships. CERIF enables quality maintenance, archiving, access, and interchange of research information (also between CRIS) and supports knowledge transfer to decision makers, for research evaluation, research managers, strategists, researchers, editors, and the general public. CERIF is recommended by the European Union to its member states.
- *CrossRef* provides the technical infrastructure for linking references between scientific and scholarly publications (journal articles, books, conference proceedings, working papers, technical reports, data sets) using digital object identifiers (DOIs). CrossRef is provided by a not-for-profit association with members from the publishing industry.
- *DataCite* provides persistent identifiers (DOIs) for research data. DataCite is also a non-profit organization with the aim to help the research community locate, identify, and cite research data with confidence.
- *ORCID* provides a persistent digital identifier for individual researchers across the world. Among other things, ORCID may solve the problem with author name disambiguation in bibliographic data. ORCID is a non-profit member organization for research organizations, publishers, funders, professional associations, and other stakeholders in research.

EuroCRIS is the organization behind one of the five resources mentioned above, CERIF. The organization deserves more attention in this chapter because it has for a long time been the main international arena for professional and research-based developments of CRIS. EuroCRIS grew out of international collaboration on technical standards for exchange of research information that was encouraged and supported by the European Commission between 1987 and 2002 [25.9]. It is now a non-profit (and at the moment non-funded) member organization, which in addition to developing and maintaining CERIF, also hosts the major conferences and meetings for the European community of CRIS specialists.

EuroCRIS has so far been more successful in connecting people than in connecting CRIS with CERIF. This might be changing at the moment, since EuroCRIS is now joining the collaborating partners in the large-scale European OpenAIRE2020 project with the more general aim of promoting open scholarship and substantially improve the discoverability and reusability of research publications and data. OpenAIRE (openaire.eu) is a European infrastructure enabling researchers to comply with the European Union requirements for open access to research results. OpenAIRE collects metadata from a variety of data sources, i. e., publication repositories, data archives, and CRIS across Europe and beyond.

So far, the OpenAIRE search engine retrieves only freely available documents on the internet without any connection to organizational information. However, expertise from EuroCRIS has now contributed with OpenAIRE Guidelines for CRIS Managers. The guidelines specify the interoperability layer between CRIS and the OpenAIRE infrastructure. The information interchange is based on CERIF as a data model, the CERIF XML exchange format, and the open archives initiative protocol for metadata harvesting (OAI-PMH) for repository interoperability. The guidelines are intended mainly for implementers and administrators of CRIS who plan to communicate research information to OpenAIRE.

In a recently funded project, the implementation of these guidelines will now be tested using the local CRIS system at Radboud University Nijmegen, which is based on CERIF. The aim is to deliver structured CRIS information to the OpenAIRE aggregation. The project also includes collaboration with commercial CRIS vendors (Sect. 25.4) to support their effort towards achieving OpenAIRE compatibility.

The OpenAIRE and CRIS project is yet another example that the technical resources for interoperability are becoming available among research information managers and on a commercial market. However, the interoperability of CRIS is also about contents and will in the end depend on several other factors. We shall end this overview of initiatives and recommendations at the international level by pointing at three other such factors.

It has been shown by *Vancauwenbergh* et al. [25.10] that adopting the CERIF standard is not sufficient for efficient and accurate exchange of research information. Providers of information may use different words for the same concept and vice versa. There will be a need for definitions and a glossary, as well as a model for governance of research information and shared classifications to operationalize CERIF [25.1, 11]. The answer to the need for a more general model could be an approach to defining and specifying CRIS data based on an ontology-based data management (OBDM) model, see *Daraio* et al. [25.12] and the chapter on OBDM in this volume.

A second, related factor is that there will be a need for solutions and decisions that go beyond technical standards and address the same main problems observed by *Daraio* and *Glänzel* [25.1] in relation to data integration for research administration, evaluation, and policy, i.e, data quality, comparability, standardization, interoperability, modularization, classifications and concordance tables, comprehensiveness and completeness, and shared schedules for updating. We will address such problems in the second half of this chapter. To take one example, the problem of comparable subject classifications—which is mentioned in the *Position statement on research information systems* by *Science Europe* [25.4]—will be discussed concretely in Sect. 25.7.

The third factor concerns policies and decision-making at all levels. There needs to be a vision of improvement of research organizations through shared information and comparability, as well as a willingness to decide and invest in shared information systems, at the policy level. There also needs to be an interaction between research organizations, their funders, and the expertise in research information, documentation, indicators, and evaluation. Locally, there needs to be bottom-up engagement in finding shared solutions. Often, only local purposes are on the horizon, resulting in different and incompatible local solutions for the same general needs. While academics usually prefer global exchange of information and data, administrators tend to have an intramural perspective on information management. Moreover, as we shall see in the next section, the commercial market for CRIS favors a model for serving mainly local needs, while the data and the technical solution stay with each commercial provider. It is still unclear whether the solution for an internationally integrated CRIS can be provided by a commercial supplier.

25.4 Commercial Solutions to CRIS

CRIS may be the answer to internal needs for more professional information management within research organizations as well as to external needs for public information and for institutional data indicators in national research evaluation and funding systems [25.13]. Both types of needs are creating a rapidly expanding market for CRIS solutions in which the four most widespread products are *Pure*, provided by Elsevier, *Converis*, provided by Clarivate Analytics, *Elements*, provided by Symplectic, and *Researchfish*, provided by the organization with the same name in Cambridge, UK.

The first three of these products are designed to serve research administration purposes within research performing organizations. They typically connect all types of individual-level data on resources, activities, and outputs from internal and external data sources. They promise *an overview of all accomplishments, with advanced analytic reports of outputs and impact* (Converis), *a single point of organization, presentation, and reporting for all scholarly and research activities* (Elements), and *an evidence-based approach to your institution's research and collaboration strategies, assessment exercises, and day-to-day business decisions* (Pure).

Researchfish provides much of the same information but is designed in response to the needs of research funding and evaluation organizations. An interesting example of an implementation of Researchfish can be inspected at the Novo Nordisk Foundation, a large private funder of research in Denmark, which uses the system to collect data for impact assessment. Their solution is interesting from a data integration point of view by enabling researchers to enter an output or outcome just once and then re-use that data in reports to multiple funders, not only the Novo Nordisk Foundation, and for their own use, such as in CVs.

Presently, the trend is towards more comprehensive commercial CRIS systems serving several purposes. As an example, Elsevier is integrating Pure with other well-established and newly acquired products, such as Mendeley, SciVal, and Scopus, to be the information provider for all purposes in research activities and in research administration, evaluation, and funding.

Interestingly, all four of the above-mentioned CRIS products were originally designed by computer and information scientists or students at university campuses (in Denmark, Germany, and the UK) before commercialization. Several universities still maintain their locally designed non-commercial CRIS solution. At the national level, non-commercial solutions are more widespread than commercial solutions. Examples of non-commercial CRIS at the national level are the R and D information system in the Czech Republic, the Estonian research information system, and the Current Research Information System in Norway (CRISTIN).

While standardized CRIS solutions are commercially marketed across countries, they remain local silos of data. As an example, Denmark has eight universities, all of which are customers of Elsevier by using local applications of the Pure system. The applications are not integrated, although this would have been practical, since all eight universities use standardized procedures to collect bibliographic records through their Pure system for the Danish bibliometric indicator at the national level, which is one of the indicators in the performance-based research funding system in Denmark. The Ministry of Higher Education and Science had to build its own national database on top of the local Pure systems.

The reason for not providing integrated solutions on the commercial side seems to be that the standardized CRIS solutions are designed and tailored for a market of individual research organizations. Each customer is guaranteed that the information will not be shared with other customers. In addition, national needs in this area are often expressed through individual organizations. As an example, the research excellence framework (REF) in the UK requires that the data for the assessment is submitted by each university without being shared between them. Thus, each university must create or purchase its own solution to data production.

From a technical point of view, commercial solutions to CRIS are generally more efficient and advanced than non-commercial solutions. However, from the point of view of integrating CRIS and sharing data in an open space, only non-commercial solutions have solved the challenges so far.

25.5 Agreeing on Sharing Well-Defined Data

In several countries, research institutions may include information in their local CRIS that they would hesitate to make publicly available, particularly if they see their own institution as competing with other institutions for staff and resources. There may even be legal restrictions and privacy issues. Still, the existence of several non-commercial national CRIS already demonstrates that such problems can be solved technically and by agreements between the institutions. However, organizational and legal solutions in CRIS are much less shared and discussed at the international level than the general technical solutions are. As we shall see here, Germany has provided an example of how these matters can be brought to the international level.

National CRIS are typically more widespread among the smaller European countries with a unified research system than they are in larger countries with a more heterogenous and distributed research system. Agreements on sharing well-defined data may be more difficult to reach in larger countries and between countries. An important source of experience and learning in this respect, covering possible obstacles and achievements as well as necessary considerations in the process, is the national research core dataset project for the science system in Germany [25.14]. Germany has regionally funded universities as well as non-university research institutions with federal funding. As a federal state, Germany has lacked nationwide reporting standards and data on research activities and outcomes. The research core dataset project has worked on defining uniform classification systems, definitions of core concepts, assessment of local adaptation costs, decisions on what information will be useful to share, and issues dealing with central steering versus local autonomy and openness versus privacy. The experiences and solutions in the German project are, thereby, directly relevant for all the issues that still need to be solved for internationally integrated CRIS—even if the technical solutions are available and the open science agenda is acknowledged everywhere.

25.6 Testing Real Data Sharing in the Social Sciences and Humanities

The *European Network for Research Evaluation in the Social Sciences and the Humanities* (ENRESSH, The European Cooperation in Science and Technology (COST) action 15137) currently brings together more than 125 experts from 36 countries with the aim of developing and proposing best practices in the field of social sciences and humanities (SSH) research evaluation. One of the work groups is focused on information systems, databases, and repositories for publications and other outcomes of SSH research with the explicit aim of *designing a roadmap for a European database*. One already published result is an investigation of possible data sources, including CRIS, for scholarly publications in books [25.15]. Two other projects within the group are particularly interesting in the context of internationally integrated CRIS.

ENRESSH and the Centre for Research and Development Monitoring (ECOOM) at the University of Antwerp recently conducted a survey across Europe and found that 23 out of 39 countries have a database covering scholarly publishing in the humanities and social sciences beyond the references indexed in WoS or Scopus [25.16]. There is, however, large variability rather than standardization in how the data is defined and structured. As an important starting point for further progress, the survey also identified the presence of a legal framework and the specific responsible organization behind each national initiative.

A second interesting project [25.17] is being conducted by ENRESSH in collaboration with the Centre for Scientific Computing (CSC) in Helsinki, with support from Finland's Ministry of Education and Culture. CSC has developed and now operates the VIRTA publication information service, which provides a national-level solution for integrating publication information in Finland. It has already proved it can solve the problem that Finnish universities record their data in different types of commercial and non-commercial CRIS solutions. A wider application of the VIRTA concept at the European level has been tested by using real institutional data from CRIS in four countries, Belgium (Flanders), Finland, Norway, and Spain [25.18]. Some of the conclusions from the VIRTA pilot study with data from four countries are:

- The pilot study demonstrates that it is possible to integrate institutional publication data from different countries using the VIRTA model. This required the identification of data fields that all participating institutions and countries could supply (the *lowest common denominator*).
- The main challenge is that institutional and national data sources use different data models as well as different data collection and validation procedures. Agreed upon data definitions and classifications will be needed to solve this problem.
- It is also possible to increase the comparability of data by developing automated methods to restructure and reclassify data from different CRIS.

25.7 Subject Classification

Subject classification of organizations or outputs may be necessary to make data from CRIS useful for analytical purposes. Accordingly, the *Position statement on research information* from *Science Europe* [25.4], which was discussed in Sect. 25.3, suggests providing full documentation on classification systems, including subject definitions, and encourages cross-mapping between them. The CRIS that we want to integrate, or exchange data between, or compare data from, may have different types of subject classification systems [25.1]:

- *Cognitive* (content-related, used in libraries, bibliographic databases, patent and trade offices)
- *Administrative* (responsibility-related, used by authorities, funding organizations)
- *Organizational* (structure-related, used by institutions according to their internal organizational structures)
- *Qualification-based* (competency-related, reflecting the skills of individuals or groups of persons.

Subject classification is one of the typical unavoidable problems that arise as soon as the process of integrating CRIS moves from the technical solutions to dealing with real content. This situation was very concretely encountered and had to be solved in the VIRTA project described above. The bibliographic data from Belgium (Flanders) had an organizational classification, while the data from Finland, Norway, and Spain had different types of cognitive classifications. The problem had been encountered and solved ad hoc in an earlier study of CRIS data from Flanders and Norway [25.19], but it needed a more forward-looking solution for the VIRTA pilot study. A concordance was created using the revised field of science and technology (FoS) classification of the Organisation for Economic Co-operation and Development (OECD) [25.20].

Still, underlying the different classification systems, there may also be different methods for assigning fields to publications. For example, in Finland and Spain, fields are assigned cognitively to publications at record level. Norway uses the same method for book publications, but assigns fields for journal articles cognitively at journal level. In Flanders, as was mentioned, publications are classified on the basis of the organizational unit. In the case of journal articles, the *Norwegian* solution with a journal-level classification was chosen, however with the more principal choice of the OECD FOS classification as the basis for the concordance. This solution could not be implemented without a standardized register of journals, which leads to the next level for defining contents in integrated CRIS, which we will discuss in this chapter.

25.8 Dynamic Registers of Evaluated Scholarly Publication Channels

In any CRIS with bibliographical records, if the bibliometric analysis is to be taken beyond simple counts within types of publications (e. g., journal articles, book chapters, monographs), a dynamic register of publication channels with standardized titles, identifiers, and subject classifications is needed. To retrieve data and compare across different CRIS, a shared standardization of the registers will be needed as well. An example of a minimum of standardized information in this type of register would be:

- Publication channel type: Journal
- Title: Scientometrics
- ISSN: 0138-9130
- OECD FOS classification: 5.8 Media and communications.

Related to the problem of subject classification of publication channels is how to define those that meet scholarly standards, for example regarding peer review practices. Ulrich's periodicals directory covers more than 300 000 periodicals worldwide, but they are of all types and include popular magazines, newspapers, and professional journals as well as academic and scholarly journals. Whenever a particular selection of journals is not sufficient, e. g., WoS core collection with 18 000 journals or Scopus with a selection of 23 000 journals, there is the need to agree on the criteria for a wider selection of academic and scholarly journals. This need has been particularly apparent in the social sciences and humanities [25.2].

In response to the situation in the social sciences and humanities, the European Reference Index for the Humanities (ERIH) was originally created and developed by European researchers under the coordination of the Standing Committee for the Humanities (SCH) of the European Science Foundation (ESF). The ERIH lists, which initially mainly covered disciplines in the humanities, were first published by ESF in 2008, while revised lists were made available in 2011–2012. In 2014, the responsibility for the maintenance and operation of ERIH was transferred to the Norwegian Centre for Research Data (NSD), a non-commercial organization owned by the Norwegian Ministry of Education and Research. NSD also runs the Norwegian register of scientific journals, series and publishers as a resource for CRISTIN. The international register of journals and series at NSD is now called ERIH PLUS to indicate that it has been extended to the social sciences.

The relevance of ERIH PLUS to CRIS is that one of the aims is to provide a well-defined, standardized, and dynamic register of scholarly journals and series in the SSH, thereby making data available and comparable across different CRIS [25.21]. Although ERIH PLUS is limited to the humanities and social sciences, its six criteria for the inclusion of journals could be a starting point for defining scholarly and scientific journals more generally:

1. Explicit procedures for external peer review
2. Academic editorial board, with members affiliated with universities or other independent research organizations
3. Valid ISSN code, confirmed by the international ISSN register
4. All original articles should be accompanied by abstracts in English and/or another international language relevant for the field
5. Information about the affiliations and addresses of the authors should be published for each article
6. National level as a minimum: No more than two-thirds of the authors published in the journal are from the same institution.

Note that the last criterion ensures that the register will not be endless and that there is a proper basis for independent peer review. Criteria 1, 2, and 6, taken together, ensure that ERIH PLUS promotes research quality in the SSH. By allowing for journals published in the national languages, societal relevance is also promoted. Criteria 3–5, taken together, ensure that data will be efficiently relevant, searchable, and comparable across CRIS and other bibliographic data sources. They are also required for providing data for bibliometric analysis.

In a project presently funded by Nordforsk, the organization that facilitates and provides funding for Nordic research cooperation and research infrastructure, Denmark, Finland, Iceland, Norway, and Sweden together are creating one merged Nordic list of scholarly publication channels as a shared resource for their national CRIS. The criteria for inclusion are in concordance with the criteria used in ERIH PLUS. The Nordic list includes all areas of research, not only the social sciences and humanities, and it will be based on the OECD FOS classification.

25.9 Ensuring Comprehensiveness of Data in a CRIS

The preceding sections of this chapter have progressively dealt more concretely with contents-related challenges that need to be solved in an internationally integrated CRIS. As we now move on to the level of a national CRIS, we encounter specific problems and possible solutions that do not surface in international overviews of the situation. These problems are related to the completeness and quality of the data. There is a need to agree on definitions, instructions, and procedures to ensure data quality. There may also be a need to incentivize the data production to achieve completeness. These challenges are not technical, but the technical design of the CRIS can be part of the solution, e. g., by facilitating easy quality assurance in the data production line.

CRISTIN (the current research information system in Norway, cristin.no), the national CRIS of Norway, is an example of a CRIS with several years of experience with meeting and solving challenges with data quality and completeness. The system was developed in 2003 at the University of Oslo and used independently of each other by the four major universities until 2010. By that time, the government had facilitated a 2 year process in which it was agreed and specified how the local CRIS could become *one shared information system* for almost all public research organizations in the country, including universities, other higher education institutions, independent research institutes, and hospitals [25.22]. The system is now provided by the Norwegian government's Directorate for ICT services in higher education and research (UNIT).

Like most CRIS, the Norwegian system has daily updated, standardized, and searchable information about researchers and their affiliations, projects, and outputs. All kinds of outputs can be registered. The costs of running the national CRIS would not be legitimate without multiple use of the same data. References to publications are registered only once, after which they can be used in CVs, applications to research councils, evaluations, annual reports, internal administration, bibliographies for Open Archives, links to full text, etc.

A part of CRISTIN is called the *Norwegian Science Index*. Only in this part can one expect to find complete, quality-assured, and well-structured data that can also be used in bibliometric analyses. The Norwegian Science Index is based on particular definitions, instructions, and procedures to ensure data quality, because it is used as the basis for one of the indicators in performance-based research funding models that affect most of the participating research organizations in the higher education, independent institute, and health sectors. The link to funding also explains why completeness can be expected. We will discuss the issue of completeness first and then turn to how data quality is ensured.

The relation between CRIS data and funding is, in this case, known as the *Norwegian model* [25.13, 23, 24], which so far has been adopted at the national level by Belgium (Flanders), Denmark, Finland, Norway, as well as at the local level by several Swedish universities and by University College Dublin. The national model has three components:

A A complete representation in a national database of structured, verifiable and validated bibliographical records of the peer-reviewed scholarly literature in all areas of research.

B A publication indicator with a system of weights that makes field-specific publishing traditions comparable across fields in the measurement of *publication points* at the level of institutions.

C A performance-based funding model that reallocates a small proportion of the annual direct institutional funding according the institutions' shares in the total of publication points.

In principle, component C is not necessary to establish components A and B. The experience, however, in all the above-mentioned countries, is that the funding models in C support the need for comprehensiveness and validation of the bibliographic data in component A. In Norway, the bibliometric indicator based on CRIS data reallocates less than 2% of the total funding, but it still stimulates a completeness, which is valuable not only for the funding instrument, but also for all other uses of CRIS data. Without the link to funding, which was introduced for the higher education sector in 2004, CRISTIN would probably not have been restructured as a *shared* national system since 2010. One of the reasons for this investment was to provide for increased transparency, more uniform quality-assurance, and shared resources in the data production line.

Component B, the publication indicator, is the first bibliometric indicator to give a balanced representation of productivity in all fields [25.25]). It does so by building on a definition of outputs that includes publications not indexed in the WoS or Scopus. It applies an intermediate solution between whole counts and fractionalization in cases of co-authored publications, using the square root of the fraction. Bibliographic data from CRISTIN can, however, be used for any type of analysis or indicator. A few examples are given in Sect. 25.11 of this chapter.

25.10 Ensuring the Quality and Consistency of Data in CRIS

The possible link to institutional funding can also be discussed in relation to quality-assurance of the data. Most countries with a national CRIS have a funding model for research that makes use of it. In these countries, the quality assurance of the data is necessary for the acceptance and legitimacy of the funding instrument. Croatia and Estonia are presently examples of countries with a national CRIS not used for funding because the data quality is so far not trusted. WoS and Scopus are used instead for the funding instrument. These examples clearly indicate that once the decision is taken to use CRIS data for funding allocation, there is also a need to implement several measures to ensure data quality and consistency. We will return to the example of the Norwegian Science Index in CRISTIN to explain how this has been done.

The data for the Norwegian Science Index in CRISTIN are delimited by a definition of scholarly publications, the development of which representatives from all areas of research contributed to and agreed on before it was implemented in 2004. According to this definition, a scholarly publication must:

1. Present new insight
2. Have a scholarly format that allows the research findings to be verified and/or used in new research activities
3. Be in a language and with a distribution that makes the publication accessible for a relevant audience of researchers
4. Be in a publication channel (journal, series, book publisher) that represents authors from several institutions and organizes independent peer review of manuscripts before publication.

While the first two requirements of the definition demand originality and scholarly format in the publication itself, the third and fourth requirements are supported by a standardized dynamic register of approved scholarly publication channels at http://dbh.nsd.uib.no/kanaler/. Suggestions for additions can be made at any time through the same web page. This means that the two last criteria have a centralized solution. It is operated by NSD, the Norwegian centre for research data, in collaboration with the National Publishing Board, which has an academic representation across all major research areas and types of research organizations. The register now not only supports the CRIS and the definition above. It is also used for information about publication channels with open access, promoting those that fulfill the definition and excluding the exceptions. This task is now organized in collaboration with similar registers in the other Scandinavian countries and the Directory of Open Access Journals (DOAJ).

It is the responsibility of each institution to apply the two first criteria. Publication channels are usually hybrid. Not all articles in *Nature*, and not all books published by *Oxford University Press*, will fulfill the first two criteria. The institutions need to judge and select correct publication types when they register data for the Norwegian Science Index. They are aided in doing so in several ways:

- By agreement with Elsevier, data for publications with Norwegian affiliations that are indexed in Scopus, including publication type, are continuously imported to CRISTIN. Thus, around 75% of the publications need only be validated by the authors and the institutions.
- Depending on the size of the institution, each of them has appointed one or more administrative contact persons with a local overall responsibility for registrations in CRISTIN. These persons monitor the registration process in collaboration with people at CRISTIN. Cases of doubt will first be discussed with researchers in leading positions at their institution and then with the CRISTIN organization if needed. Cases of doubt will also be returned from CRISTIN after the annual registration process in April, when the institutions take formal responsibility for the data produced in the previous year.
- A document is provided with guidelines for registering research publications in the Norwegian Science Index. The guidelines present further and more detailed explanations with examples of the application of the definition above.

The methods for quality assurance and the incentives for completeness require resources as well as organizations with clear responsibilities. Guidelines for registration are often discussed among countries with national CRIS (e. g., Denmark, Finland, Flanders, Norway), and the countries inspire each other in their further development. The most elaborated guidelines presently that are also available in English, can be downloaded from the *Danish Ministry of Higher Education and Science* [25.26]. The guidelines include:

- An operational definition of research publications (similar to the Norwegian definition given above) with clarifications of each of the criteria, including examples and more detailed definitions, e. g., of peer review.
- Requirements for the registration of the authors of a publication, e. g., all authors must be registered in the same order as they appear in the publication.
- The criteria for inclusion in the register of publication channels.
- Descriptions of publication types to be included, and those not to be included, with explanations and examples of each type. This section is the largest part of the guidelines.

As mentioned above, a survey to 39 countries found large variability rather than standardization in how the data are defined and structured in national CRIS or bibliographic databases [25.16]. The Danish guidelines discussed here are closely related to an implementation of the *Norwegian model* (Sect. 25.9) and not typical of most CRIS. However, the problem with defining scholarly and scientific publications can be found in any CRIS that try to be interoperable with Scopus or WoS, which is exactly what the widespread commercial solutions to CRIS try to do.

25.11 Examples of Studies of Research Based on CRIS Data

The establishment of non-commercial national CRIS in some countries has supported an increase in output-based studies of research, particularly in the humanities and social sciences, where CRIS can provide a more complete representation of scholarly publications than we find in commercial data sources [25.27]. The field is quite new, with most of the publications having appeared only the last 2 years, after the establishment of the COST network ENRESSH mentioned above. The increased activity is now the basis for a new series of biannual conferences focusing on research evaluation in the social sciences and humanities (RESSH, established in Rennes in 2015 and continued in Antwerp in 2017).

The ECOOM group at the University of Antwerp is particularly active. This group developed and runs the Flemish bibliographic database for the social sciences and humanities (Vlaams Academisch Bibliografisch Bestand voor de Sociale en Humane Wetenschappen VABB-SHW) for a similar purpose to the Norwegian CRISTIN system, see Sect. 25.9. They use the data for studying several aspects of the publishing patterns of the social sciences and humanities that have rarely been studied before. Here are some examples:

The ECCOM group studied general changes in the publication patterns of the social sciences and humanities over a decade (2000–2009), finding growth in the output, particularly a steady increase in the number and proportion of publications in English, however with no overall shift away from book publishing [25.28]. They found almost identical evolutions in the use of English as a publication language by comparing data from CRIS in Flanders and Norway, however WoS coverage was stable for Norway but had been increasing rapidly for Flanders, probably because of differences in the parameters used for performance-based funding of universities [25.19]. Internationalization was also found in book publishing. Whereas peer reviewed books were increasingly published abroad and in English, non-peer reviewed book literature remained firmly domestic and published in the Dutch language [25.29]. Whereas the humanities are more continentally oriented in their book publishing, the social sciences are firmly Anglo-Saxon oriented [25.30]. A study of co-authorship patterns in the social sciences and humanities indicated that collaborative publishing in the SSH is increasing with a sharp decline in single-author publishing [25.31]. A study of 753 peer reviewed edited books and the 12 913 chapters published therein revealed that not only co-authorships, but also co-editing and publishing different chapters in the same books are indicators of scholarly collaboration in the social sciences and humanities [25.32]. The editors of scholarly books are mostly established researchers, produce more book chapters and monographs than do other researchers, and are more productive [25.33].

A new study based on CRIS data [25.34] investigates publication patterns in the language and type of social sciences and humanities across a much wider range of non-English speaking European countries, including Eastern Europe, i. e., Czech Republic, Denmark, Finland, Flanders (Belgium), Norway, Poland, Slovakia, and Slovenia. The study demonstrates that publication patterns are related not only to discipline but also to each country's cultural and historic heritage. This finding corrects an assertion in an earlier CRIS-based study [25.27] that publication patterns vary by discipline, but less across countries in the same discipline.

Also on the basis of CRIS data, other researchers have provided deeper insight into the publishing patterns of particular fields of research, such as political science [25.35, 36] and law [25.37].

There are some studies based on CRIS data that investigate policy-related questions across all fields of research, not only the social sciences and humanities. With data from the CRIS of the University of Helsinki, *Puuska* [25.38] examined the effects of a scholar's position and gender on publishing productivity in several types of scientific publications, such as monographs, articles in journals, articles in edited books, and articles in conference proceedings. *Aksnes* et al. [25.39] studied the mobility of researchers on the basis of CRIS data from the four main Norwegian universities.

Other studies have contributed to a critical examination of how CRIS data are used for statistics, evaluation, and funding in research management, most often with suggestions for further development of data and indicators [25.15, 40–44], sometimes only describing potentially negative effects of such use [25.45].

Finally, to illustrate in more detail that the use of CRIS data in studies of research may also have a broader interest beyond bibliometrics and studies of the social sciences and humanities in particular, we will end this chapter by presenting two examples showing that bibliographic data in CRIS can be combined with other data (personal variables, backgrounds, resources, projects, networks, memberships) within or outside of the CRIS system, thereby serving science studies more generally.

The first example is a little study of gender, age, and productivity that we did some years ago, based on data in the Norwegian CRISTIN system [25.46]. Here, gender, age, and complete records of all peer reviewed scientific publications are among the available information for each active researcher. We studied the productivity of 17 212 researchers (10 279 men and 6933 women) aged 27–67 who published in 2011. Altogether, they contributed to 12 441 unique publications. There was no double counting if two or more researchers contributed to the same publication. Instead, publications with multi-authorship were fractionalized by the number of authors. Figure 25.1 shows the result by presenting the women's share among Norwegian researchers and their publication output in each 1 year age cohort between 27 and 67.

We can see the gender gap decreasing as younger generations are recruited to research. We also observe that the difference in productivity between men and

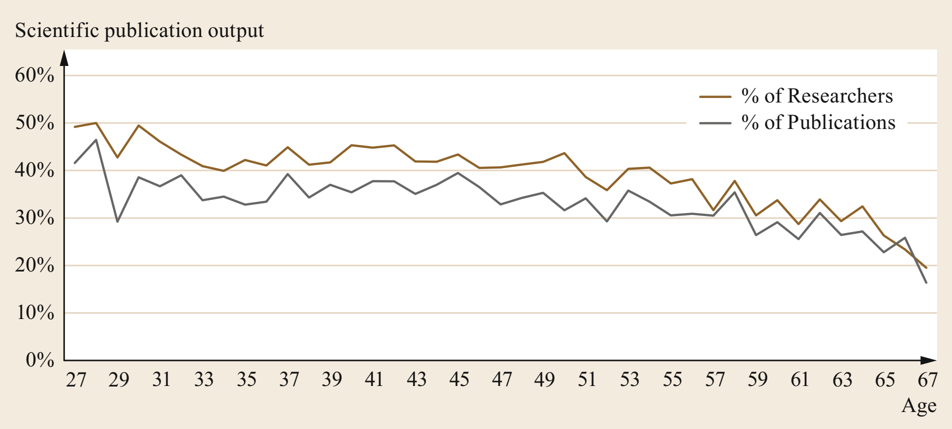


Fig. 25.1 Age and women's share of Norway's researchers and their total scientific publication output in 2011. Based on data from CRISTIN, representing more than 17 000 active researchers working at 160 different research institutions in Norway

women is somewhat larger in the younger age cohorts. This is not a new finding. The same observation and its possible explanations were studied more extensively in previous studies, e. g., by *Kyvik* and *Teigen* [25.47]) with the telling title *Child care, research collaboration, and gender differences in scientific productivity*. That study, however, was based on a survey and interviews with relatively few researchers. Figure 25.1 is based on complete data for all active researchers in a country. With the help of the CRIS system, we can see that the difference in productivity between men and women is so far consistent across all types of institutions (universities, university colleges, research institutes, hospitals) and across all fields of research (humanities, social sciences, health sciences, and natural sciences). This could be an indication that gender equality in research is dependent also on the degree to which gender equality has been achieved in society.

The last example is a research project that was presented at the 16th Conference of the International Society of Scientometrics and Informetrics (ISSI) in 2017 [25.48]. The project is a response to two independent studies that were first presented at ISSI 2015 and published the year after in PLOS ONE. *Larivière* and *Costas* [25.49] and *Sandström* and *van den Besselaar* [25.50] observe similarly that productivity among individual researchers is correlated with citation impact in large datasets from WoS. While the latter study draws the policy implication that productivity should be incentivized, the first study explains their findings by the Mertonian theory of cumulative advantages and maintains that research assessment should be qualitative and focus on research quality. Both studies are based on author name disambiguation in WoS data.

As acknowledged by the authors, there are several problems with studying individual productivity by using author name disambiguation. Here, CRIS can come to aid. In our study of the same general research question, we match WoS with records in The Norwegian Science Index and The Norwegian Research Personnel Register. Hence, we can study real persons, not just authors. We know their age, gender, position and affiliation, as well as their former career and educational background in the higher education sector. Productivity depends on what roles researchers actually take in research, what positions they have, what resources are available, and what they achieve in their careers. We also have a broader basis for measuring productivity across fields, building on the completeness of CRIS data. We found that productivity and citation impact is much less correlated if publications beyond WoS are also included in the measurement of productivity. We also found higher average citation rates among post docs than among professors. Young researchers are, on the average, more cited than seniors, if measured per publication. On the other hand, senior researchers and professors are more productive within WoS.

25.12 Conclusions

Well integrated and structured current research information systems on the institutional or national levels serve several purposes at all levels, from the need for the individual researcher to record and provide information about research activities and achievements in internal and external contexts, to the need for research information, statistics, and indicators at the institutional or national level. With sufficient data quality and completeness, they are promising also as data sources for studies of researchers and their activities, including bibliometric studies. The strength of these systems is related to the completeness of bibliographical records, the automatic disambiguation of authors/persons and addresses/affiliations, and the possibility of thereby connecting with other data describing the researchers, their institutions and resources, and the outcomes of their research.

CRIS would become even more interesting for studies of research if such systems were standardized and interoperable for exchanging comparable data on the international level. International integration of CRIS is already on the European policy agenda for several other reasons. This chapter has focused on factors that will determine the success of CRIS integration. Generally, *technical* solutions to different aspects of an integrated CRIS are already available. The challenge is instead to coordinate the processes and agreements that are needed before *contents* are well defined and can be efficiently shared in an integrated CRIS. The chapter has presented documents, projects, and CRIS implementations that concretely deal with the major challenges and represent solutions that can contribute to further advancement. It also presented examples of how CRIS-based data may be used in policy-relevant research that can enlighten questions that are otherwise not easy to approach.

One might ask why European-level support to developing technical solutions to CRIS integration has been successful only in creating the solutions themselves, not in actually integrating the systems or making them exchange information. The CERIF data model was developed with European support already 30 years ago, but has rarely been implemented. CERIF now has a new chance within the large-scale OpenAIRE project, which has mainly an open access agenda and will probably not solve what *Daraio* and *Glänzel* [25.1] list as the main problems in relation to data integration for research administration, evaluation, and policy as data quality, comparability, standardization, interoperability, modularization, classifications and concordance tables, comprehensiveness and completeness, and shared schedules for updating.

This chapter has described how such problems can be dealt with concretely. However, except for the bottom-up initiatives that we mentioned within the ENRESSH COST network and the Nordforsk network of Nordic countries, there is no follow-up of initiatives to address these problems at the international level. Just as the *European Scoping Project* [25.6] had no practical outcome, the *European Report on Measuring Scientific Performance for Improved Policy Making* [25.3] and *Science Europe*'s [25.4] *Position Statement on Research Information Systems* have so far not brought Europe many steps closer to CRIS integration.

We explain this somewhat depressing situation by the fact that there are three competing models that seem to be able to achieve the same goal—without being successful either. Including the model discussed in this chapter, we have the following four so far unsuccessful models:

- *The informatics model*, with CERIF and OpenAIRE given above as examples
- *The survey model*, with the European Tertiary Education Register (ETER) and U-Map/U-Multirank as examples
- *The commercial model*, with Converis, Elements, and Pure as examples
- *The open CRIS model*, with the VIRTA project presented in Sect. 25.6 as an example.

The strengths of the survey model and the commercial model is that they ensure that institutions can retain information in their local CRIS that they hesitate to share with other institutions or with the national level. The survey model is based on an agreed selection of items and indicators to provide information for. However, as shown in Sect. 25.5 with the research core dataset project in Germany, it would be possible to establish such agreements with an open CRIS model as well. It is technically feasible to share only parts of the information available in a local CRIS.

The additional strength of the commercial model is that it allows for tailor-sewed local CRIS applications of a technically very advanced general product that also exchanges information with other commercial and open information sources. These products immediately serve local research management with useful tools for strategic decisions and development. It may be seem to be an additional strength that the local information is not shared with other organizations except for the commercial provider, which in turn creates benchmarking information based on local information from all of its customers.

Hence, according to the commercial model, the total overview remains with the commercial provider and is not shared in the public space. This situation is satisfying if research organizations see themselves more as competitors than as partners within academia, which is increasingly the case. To counteract this trend and support the open CRIS model, there needs to be a vision of improvement of research organizations through shared information and comparability, as well as a willingness to decide and invest in shared information systems, at the policy level. There also needs to be an interaction between research organizations, their funders, and the expertise in research information, documentation, indicators, and evaluation.

References

25.1 C. Daraio, W. Glänzel: Grand challenges in data integration—State of the art and future perspectives: An introduction, Scientometrics **108**(1), 391–400 (2016)

25.2 G. Sivertsen, B. Larsen: Comprehensive bibliographic coverage of the social sciences and humanities in a citation index: An empirical analysis of the potential, Scientometrics **91**(2), 567–575 (2012)

25.3 B. Mahieu, E. Arnold, P. Kolarz: *Measuring scientific performance for improved policy making* (European Parliamentary Research Service, Brussels 2014)

25.4 Science Europe: Position Statement on Research Information Systems. Co-ordination by Science Europe Working Group on Research Policy and Programme Evaluation. D/2016/13.324/11, http://www.scienceeurope.org/wp-content/uploads/2016/11/SE_PositionStatement_RIS_WEB.pdf (Brussels 2016)

25.5 VIRTA Research Information Management Services in Finland: https://research.csc.fi/research-information-management-services#virta

25.6 G. Lauer: The ESF scoping project 'Towards a bibliometric database for the social sciences and humanities'. In: *Research Assessment in the Humanities. Towards Criteria and Procedures*, ed. by M. Ochsner, S.E. Hug, H.D. Daniel (Springer, Zürich 2016) pp. 73–77

25.7 W.G. Glänzel, R. Beck, K. Milzow, S. Slipersaeter, G. Toth, M. Kolodziejski, P.S. Chi: Data collection and use in research funding and performing organisations. General outlines and first results of a project launched by Science Europe, Scientometrics **106**(2), 825–835 (2016)

25.8 M.D. Wilkinson: The FAIR Guiding Principles for scientific data management and stewardship, Sci. Data **3**, 160018 (2016)

25.9 B. Jörg: Übersicht Systeme in Europa. In: *Forschungsinformation in Deutschland: Anforderungen, Stand und Nutzen existierender Forschungsinformationssysteme—Workshop Forschungs informationssysteme 2011*, ed. by S. Bittner, S. Hornbostel, F. Scholze (Institut für Forschungsinformation und Qualitätssicherung, Berlin 2012) pp. 103–114, iFQ-Working Paper, No. 10

25.10 S. Vancauwenbergh, P. de Leenheer, G. van Grootel: On research information and classification governance in an inter-organizational context: The flanders research information space, Scientometrics **108**(1), 425–439 (2016)

25.11 S. Vancauwenbergh: Governance of research information and classifications, key assets to interoperability of CRIS systems in inter-organizational contexts, Procedia Comp. Sci. **106**, 335–342 (2017)

25.12 C. Daraio, M. Lenzerini, C. Leporelli, P. Naggar, A. Bonaccorsi, A. Bartolucci: The advantages of an ontology-based data management approach: Openness, interoperability and data quality, Scientometrics **106**(1), 857–871 (2016)

25.13 G. Sivertsen: Publication-based funding: The Norwegian model. In: *Research Assessment in the Humanities. Towards Criteria and Procedures*, ed. by M. Ochsner, S.E. Hug, H.D. Daniel (Springer, Zürich 2016) pp. 79–90

25.14 S. Biesenbender, S. Hornbostel: The research core dataset for the german science system: Developing standards for an integrated management of research information, Scientometrics **108**(1), 401–412 (2016)

25.15 E. Giménez-Toledo, J. Mañana-Rodríguez, T.C.E. Engels, P. Ingwersen, J. Pölönen, G. Sivertsen, F.T. Verleysen, A.A. Zuccala: Taking scholarly books into account: Current developments in five European countries, Scientometrics **107**(2), 685–699 (2016)

25.16 L. Sīle, J. Pölönen, G. Sivertsen, R. Guns, T.C.E. Engels, P. Arefiev, M. Dušková, L. Faurbæk, A. Holl, E. Kulczycki, B. Macan, G. Nelhans, M. Petr, M. Pisk, S. Soós, J. Stojanovska, A. Stone, J. Šušol, R. Teitelbaum: Comprehensiveness of mational bibliographic databases for social sciences and humanities: Findings from a European survey, Res. Eval. (2018), https://doi.org/10.1093/reseval/rvy016

25.17 H.-M. Puuska, J. Polonen, T.C.E. Engels, G. Sivertsen: Towards the integration of European research information, http://dspacecris.eurocris.org/handle/11366/593 (2017) Paper presented at the EuroCRIS Membership Meeting 2017—Spring (Dublin)

25.18 H.-M. Puuska, R. Guns, J. Pölönen, G. Sivertsen, J. Mañana-Rodríguez, T.C.E. Engels: *Proof of Concept of a European Database for Social Sciences and Humanities Publications: Description of the VIRTA-ENRESSH Pilot* (CSC and ENRESSH, Helsinki 2018), https://doi.org/10.6084/m9.figshare.5993506

25.19 T.L.B. Ossenblok, T.C. Engels, G. Sivertsen: The representation of the social sciences and humanities in the web of science—A comparison of publication

patterns and incentive structures in Flanders and Norway (2005–9), Res. Eval. **21**(4), 280–290 (2012)
25.20 OECD: *Revised field of science and technology (FoS) classification in the frascati manual. DSTI/EAS/STP/NESTI(2006)19/FINAL* (OECD, Paris 2007), http://www.oecd.org/sti/inno/38235147.pdf
25.21 G.A.V. Lavik, G. Sivertsen: ERIH PLUS—Making the SSH visible, searchable and available, Procedia Comp. Sci. **106**, 61–65 (2017)
25.22 G. Sivertsen (Ed.): Norsk vitenskapsindeks: Forslag til en felles database for vitenskapelig publisering innenfor et nasjonalt system for forskningsinformasjon, Innstilling fra en arbeidsgruppe oppnevnt av Kunnskapsdepartementet, NIFU STEP rapport 33/2008 (2008)
25.23 J.W. Schneider: An outline of the bibliometric indicator used for performance-based funding of research institutions in Norway, Eur. Political Sci. **8**(3), 364–378 (2009)
25.24 P. Ahlgren, C. Colliander, O. Persson: Field normalized rates, field normalized journal impact and Norwegian weights for allocation of university research funds, Scientometrics **92**(3), 767–780 (2012)
25.25 G. Sivertsen: A bibliometric indicator with a balanced representation of all fields. In: *Proc. 21st Int. Conf. Sci. Technol. Indicators*, ed. by I. Ràfols, J. Mollas-Gallart, E. Castro-Martínez, R. Woolley (Editorial Universitat Politècnica de València, Valencia 2016) pp. 910–914
25.26 Danish Ministry of Higher Education and Science: Guidelines for registering research for the Danish bibliometric research indicator, https://ufm.dk/forskning-og-innovation/statistik-og-analyser/den-bibliometriske-forskningsindikator/BFIs-regler/guidelines_for_registering_research_for_the_danish_bibliometric_research_indicator.pdf (2018)
25.27 G. Sivertsen: Patterns of internationalization and criteria for research assessment in the social sciences and humanities, Scientometrics **107**(2), 357–368 (2016)
25.28 T.C.E. Engels, T.L.B. Ossenblok, E.H.J. Spruyt: Changing publication patterns in the social sciences and humanities 2000–2009, Scientometrics **93**(2), 373–390 (2012)
25.29 F.T. Verleysen, T.C. Engels: Internationalization of peer reviewed and non-peer reviewed book publications in the social sciences and humanities, Scientometrics **101**(2), 1431–1444 (2014)
25.30 F.T. Verleysen, T.C. Engels: Barycenter representation of book publishing internationalization in the social sciences and humanities, J. Informetr. **8**(1), 234–240 (2014)
25.31 T.L. Ossenblok, F.T. Verleysen, T.C. Engels: Coauthorship of journal articles and book chapters in the social sciences and humanities (2000–2010), J. Assoc. Inf. Sci. Technol. **65**(5), 882–897 (2014)
25.32 T.L. Ossenblok, T.C. Engels: Edited books in the social sciences and humanities: Characteristics and collaboration analysis, Scientometrics **104**(1), 219–237 (2015)
25.33 T.L. Ossenblok, R. Guns, M. Thelwall: Book editors in the social sciences and humanities: An analysis of publication and collaboration patterns of established researchers in Flanders, Learn. Publ. **28**(4), 261–273 (2015)
25.34 E. Kulczycki, T.C.E. Engels, J. Pölönen, K. Bruun, M. Duskova, R. Guns, R. Nowotniak, M. Petr, G. Sivertsen, A.I. Starcic, A. Zuccala: Publication patterns in the social sciences and humanities: Evidence from eight European countries, Scientometrics (2018), https://doi.org/10.1007/s11192-018-2711-0
25.35 P. Chi: Which role do non-source items play in the social science? A case study in political science in Germany, Scientometrics **101**(2), 1195–1213 (2014)
25.36 P. Chi: Changing publication and citation patterns in political science in Germany, Scientometrics **105**(3), 1833–1848 (2015)
25.37 T.N. van Leeuwen, E. van Wijk, P.F. Wouters: Bibliometric analysis of output and impact based on CRIS data: A case study on the registered output of a Dutch university, Scientometrics **106**(1), 1–16 (2016)
25.38 H.-M. Puuska: Effects of scholar's gender and professional position on publishing productivity in different publication types, analysis a Finnish University, Scientometrics **82**(2), 419–437 (2009)
25.39 D.W. Aksnes, K. Rorstad, F. Piro, G. Sivertsen: Are mobile researchers more productive and cited than non-mobile researchers? A large-scale study of Norwegian scientists, Res. Eval. **22**(4), 215–223 (2013)
25.40 G. Sivertsen, T. van Leeuwen: *Science, Technology and Innovation Indicators 2014. Scholarly publication patterns in the social sciences and humanities and their relationship with research assessment* (Dialogic, Utrecht 2014), Thematic Paper 2
25.41 A.A. Díaz-Faes, M. Bordons, T.N. Leeuwen: Integrating metrics to measure research performance in social sciences and humanities: The case of the Spanish CSIC, Res. Eval. **25**(4), 451–460 (2016)
25.42 E. Kulczycki: Assessing publications through a bibliometric indicator: The case of comprehensive evaluation of scientific units in Poland, Res. Eval. **26**(1), 41–52 (2017)
25.43 M. Savic, M. Ivanovic, B.D. Surla: Analysis of intra-institutional research collaboration: A case of a Serbian faculty of sciences, Scientometrics **110**(1), 195–216 (2017)
25.44 E. Giménez-Toledo, J. Mañana-Rodríguez, G. Sivertsen: Scholarly book publishing: Its information sources for evaluation in the social sciences and humanities, Res. Eval. **26**(2), 91–101 (2017)
25.45 B. Hammarfelt, S.D. de Rijcke: Accountability in context: Effects of research evaluation systems on publication practices, disciplinary norms, and individual working routines in the Faculty of Arts at Uppsala University, Res. Eval. **24**(1), 63–77 (2014)
25.46 G. Sivertsen: Data integration in Scandinavia, Scientometrics **106**(2), 849–855 (2016)
25.47 S. Kyvik, M. Teigen: Child care, research collaboration, and gender differences in scientific productivity, Sci. Technol. Hum. Values **21**(1), 54–71 (1996)

25.48 L. Zhang, G. Sivertsen: Productivity versus citation impact: A study of persons, not just authors. In: *Proc. 16th Int. Conf. Scientom. Informetr., Wuhan*, Vol. 2017 (2017) pp. 970–975, http://www.issi2017.org

25.49 V. Larivière, R. Costas: How many is too many? On the relationship between research productivity and impact, PLoS ONE **11**(9), e0162709 (2016)

25.50 U. Sandström, P. van den Besselaar: Quantity and/or quality? The importance of publishing many papers, PLoS ONE **11**(11), e0166149 (2016)

Gunnar Sivertsen

NIFU – Nordic Institute for Studies of Innovation, Research and Education
Oslo, Norway
gunnar.sivertsen@nifu.no

Gunnar Sivertsen is Research Professor and Head of Bibliometric Research at the Nordic Institute for Studies in Innovation, Research and Education (NIFU) in Oslo. His expertise is in policy-oriented studies of research related to statistics, performance indicators, evaluation, funding, and science policy. Sivertsen has advised the development of new systems for research assessment and monitoring in the Czech Republic, Denmark, Finland, Flanders (Belgium), Norway and Sweden.

New Indi

Part D

Part D New Indicators for Research Assessment

26. Social Media Metrics for New Research Evaluation

Paul Wouters, Zohreh Zahedi, Rodrigo Costas

This chapter approaches, from both a theoretical and practical perspective, the most important principles and conceptual frameworks that can be considered in the application of social media metrics for scientific evaluation. We propose conceptually valid uses for social media metrics in research evaluation. The chapter discusses frameworks and uses of these metrics as well as principles and recommendations for the consideration and application of current (and potentially new) metrics in research evaluation.

26.1 Social Media Metrics and Altmetrics

Since the publication of the Altmetrics Manifesto in 2010 [26.1], interest in alternative measures of research performance has grown. This is partly fueled by the problems encountered in both peer review and indicator-based assessments, and partly by the easy availability of novel types of digital data on publication and communication behavior of researchers and scholars. In this chapter, we review the state of the art with respect to these new *altmetrics* data and indicators in the context of the evaluation of scientific and scholarly performance.

This chapter brings together three different strands of literature:

1. The development of principles for good and responsible use of metrics in research assessments and post-publication evaluations
2. The technical literature on altmetrics and social media metrics
3. The literature about the conceptual meaning of social media metrics.

The field of altmetrics has grown impressively since its inception in 2010. We now have regular altmetrics conferences where academic and commercial data analysts and providers meet. A number of nonprofit and for-profit platforms provide altmetrics data, and some

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_26

summarize these data in visually appealing statistical presentations. Some of the resulting altmetric indicators are now even incorporated in traditional citation indexes and are published on journal websites.

Notwithstanding this resounding success, we come to the conclusion that the term *altmetrics* is a misnomer and is best abandoned. Based on the published research since 2010, we have to conclude that there is no theoretical foundation or empirical finding justifying the lumping together of such various measures under the same term. We therefore propose to disaggregate the various data sets and indicators, in their use in research evaluation, in their conceptual interpretation and, last but not least, in their names. Many data and indicators (we use the term *metrics* to denote both *data* and *indicators*) that make up the altmetrics universe are actually data about social media use, reception, and impact. We suggest that it would be wiser to adopt the term *social media metrics* for these data and indicators, following a suggestion by *Haustein* et al. [26.2]. However, this is also not an umbrella term that can be used for all data and indicators currently denoted as altmetrics. As *Haustein* et al. [26.2] also indicate, some of these novel metrics are essentially web-based forms of traditional library data. And some data, such as Mendeley readerships, can be seen as a hybrid between bibliometric and social media data. Nevertheless, we think that introducing the term *social media metrics* would be helpful for understanding a large part of what is now simply labeled as *altmetrics*. We hope that this will stimulate the more accurate labeling of the remaining data and indicators. In this chapter, we will therefore use the term *social media metrics* whenever we refer to data and indicators about social media use, reception, and impact. We will restrict the term *altmetrics* to historically accurate references, since the term has been quite popular since 2010, and we do not want to rewrite history from the present.

The chapter is organized in six sections. The next, second, section explores the recent history starting with the Altmetrics Manifesto and puts this in the context of critiques of the traditional forms of research evaluation. The section shows the development of guidelines and principles in response to these critiques and mentions the concept of *responsible metrics* as one of the outcomes. The third section gives an overview of the currently available social media tools according to the data sources and discusses how they can characterize types of interactions as well as users. The fourth section zooms in on issues and actual applications of social media metrics. It reviews the technical characteristics of these data and indicators from the perspective of their use, the research questions that they can address, and principles for their use in evaluative contexts. In this section, we also spell out why the distinction between *descriptive* and *comparative* metrics may be useful. The fifth section discusses possible future developments including novel approaches to the problem of research evaluation itself. The sixth and last section details the limitations of the chapter and specifically mentions the need for more research on the use and sharing of data in the context of research evaluation. We end with the bibliography, which we hope will be especially useful for students and beginning researchers as well as for practitioners in the field of research evaluation.

26.2 Research Evaluation: Principles, Frameworks, and Challenges

26.2.1 Origins: The Altmetrics Manifesto

Altmetrics were introduced with the aim, among others, of improving the information used in research evaluations and formal assessments by providing an alternative to traditional performance assessment information. The Altmetrics Manifesto called for new approaches to fully explore the potential of the web in scientific research, information filtering and assessments. It characterized peer review as *beginning to show its age*, since it is *slow, encourages conventionality, and fails to hold reviewers accountable*. Citations, on the other hand, are *useful but not sufficient*. Some indicators such as the h-index are even slower than peer-review, and citations are narrow, neglect impact outside the academy and ignore the context of citation. The journal impact factor, which was identified by the manifesto as the third main information filter, is often incorrectly used to assess the impact of individual articles, and its nature makes significant gaming relatively easy. Since new uses of the web for data sharing and scholarly publishing have created new digital traces, these could be harvested and converted to new indicators to support researchers in locating relevant information as well as in evaluating the quality or influence of scientific work.

The idea that the web would lead to novel markers of quality or impact was in itself not new. It had already been identified by scientometricians in the 1990s [26.3–5]. This did not immediately change evaluative metrics, however, because data collection was difficult and the web was still in its early stages [26.6, 7]. Only after the development of more advanced algorithms by computer scientists did social media metrics turn into a real-world alternative in the area of scientometrics and research evaluation [26.8].

The emergence of social media metrics can thus be seen as motivated by, and contributing to, the need for responsible metrics. Its agenda included the study of the social dimensions of the new tools while further refining and developing them. Possible perverse or negative effects of the new indicators were recognized, but they were not seen as a reason to abstain from innovation in research metrics [26.8]. Experts in webometrics and scientometrics tended to be a bit more wary of a possible repetition of failures that had occurred in traditional scientometrics [26.9, 10]. As a result, the development of tools like the *Altmetric donut* did not completely satisfy the need for guidelines for proper metrics in the context of research evaluation, although they did open new possibilities for measuring the process and outcome of scientific research.

26.2.2 Standards, Critiques and Guidelines

This lacuna was filled by two somewhat independent developments. From the altmetrics community, an initiative was taken to develop standards for altmetrics indicators and use in the context of the US National Information Standards Organization (NISO) as a result of a breakout session at the altmetrics12 conference (http://altmetrics.org/altmetrics12) [26.11]. In parallel, guidelines were developed as a joint effort among researchers responsible for leading research institutions, research directors and managers, metrics and evaluation experts, and science policy researchers [26.12]. They mainly developed as a critique of the increased reliance on various forms of metrics in post-publication assessments, as in the *San Francisco Declaration on Research Assessment* (DORA) and the *Leiden Manifesto for research metrics* [26.13, 14]. It should be noted that these initiatives did not come out of the blue, but built upon a long trajectory in which the scientometric community had developed methodological standards and common interpretations of what the various indicators represent in the context of research evaluation. It led to a set of professional standards, some of them explicit, others more implicit, that guided the work of the most important metric centers [26.15, 16]. In general, the scientometric community had developed a consensus about the need to use bibliometrics as a complement to, rather than replacement of, peer review, which is summarized in the concept of *informed peer review*.

With the rise of the web and the wider availability of both traditional and novel metrics, the scientometric professionals lost their monopoly, and what was variously called *amateur scientometrics* or *citizen scientometrics* started to take off [26.15, 17–19]. This required a new approach and a more explicit nontechnical development of guidelines, for which the groundwork was laid at a series of conferences in the years 2013–2016 and in the context of the debates about the role of metrics in national research assessments, especially in Northwestern Europe.

The *San Francisco Declaration on Research Assessment* (DORA) [26.14] made 18 recommendations aimed at scholars, funders, institutions and research metrics providers. The most important recommendation was to avoid using the journal impact factor to judge the merit of individual articles or authors. Instead, article-level metrics were recommended. It also emphasized the value of all scientific outputs including data sets and software in addition to research publications. Openness regarding criteria in assessments and transparency of data and indicators is also an important theme in the recommendations.

26.2.3 Individual-Level Metrics

At the 2013 conference of the International Society for Scientometrics and Informetrics (July 2013, Vienna) and the 2013 Science and Technology Indicators/European Network of Indicator Designers (ENID) conference (September 2013, Berlin), another set of recommendations was discussed, specifically aimed at the use of indicators to assess the contribution of individual researchers [26.19].

A year later, the EU-funded project ACUMEN resulted in a more detailed evaluation guideline for both researchers and evaluators [26.20]. The core component is the *ACUMEN Portfolio*, which consists of several *pillars of evidence* (Fig. 26.1).

The basic idea of the ACUMEN approach is that evaluation is a form of communication in which the researcher herself should have a strong voice (and not only play the role of object of evaluation). The *career narrative* should be the main input for the assessment at the individual level, and qualitative and quantitative indicators can provide evidence for particular elements in the narrative. This supporting evidence is organized in three pillars:

1. Expertise
2. Output
3. Influence which enables a more flexible and modular approach to the indicators that may be used.

An important component of the ACUMEN Portfolio is the evaluation guidelines, which entail detailed advice on the merits of particular indicators covering

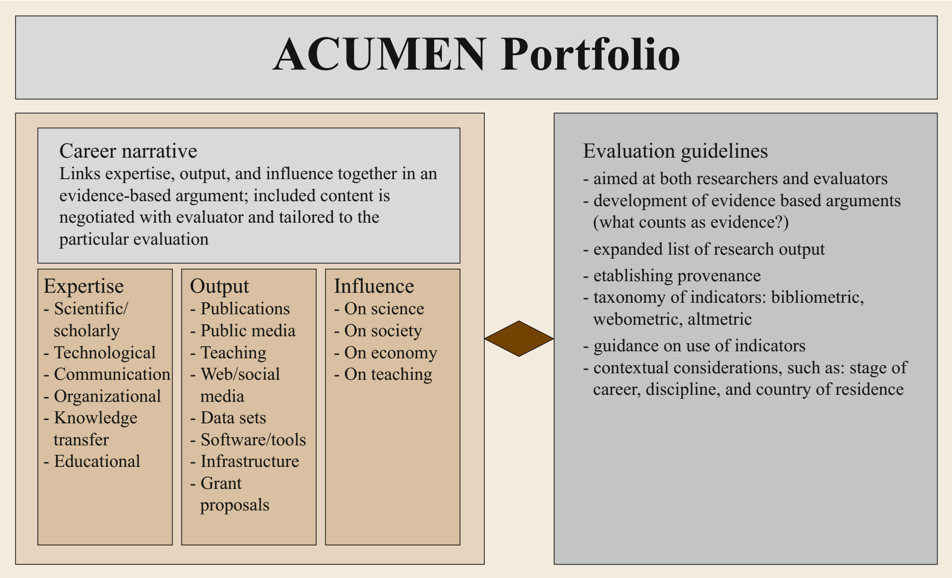


Fig. 26.1 The ACUMEN portfolio [26.20]

both traditional and alternative metrics. The guidelines are specifically aimed at both researchers under assessment and the evaluators, providing an extra layer of transparency. They are also based on the fact that researchers need to perform both roles.

26.2.4 Responsible Metrics

The Leiden Manifesto for research metrics was the result of the continuing discussion in the community of indicator specialists and scientometricians. They drew the conclusion that a public response in nontechnical terms was useful to counteract the spreading of badly used indicators in research evaluations [26.13]. The manifesto provides 10 principles that should be taken into account when using metrics in research assessment. These principles are not tied to a particular data set or assessment type. Currently, 18 translations of the manifesto have been published, which may be an indication of the need for this type of guidelines and information. Nevertheless, this does not prove that the principles are actually affecting research evaluation practices, since we may also witness symbolic adherence without changing the criteria for research evaluations or career judgments.

An even more generic framework to guide the application of quantitative indicators was laid down in the UK report *The Metric Tide* [26.12]. This was written at the request of the Higher Education Funding Council for England (HEFCE) to inform the debate about a possible replacement of the national research assessment process (which is based mainly on a massive peer review operation by panels of experts) by a metrics-based assessment. The report is not the product of one specific community, but the result of a truly interdisciplinary exercise in which researchers from a variety of fields worked together with indicators and policy experts. The report proposed to put central the concept of *responsible metrics*, echoing the notion of *responsible research and innovation* from the European science policy discourse.

The notion of responsible metrics, together with the empirical research reported in the *Metric Tide*, leads to 20 recommendations to all stakeholders in the UK research system. These recommendations support both DORA and the Leiden Manifesto and emphasize the need to put indicators in context. The research community is advised to "develop a more sophisticated and nuanced approach to the contribution and limitations of quantitative indicators". Transparency is also an important theme, with regard to both data and processes, and this should lead to a much improved research data infrastructure. The latter still lacks crucial components, especially in the area of indicators of the research environment, scientific instruments, and technical and institutional infrastructure. The *Metric Tide* pays special attention to altmetrics, with the question of whether they can complement traditional performance indicators. The overall conclusion is that current altmetrics cannot yet be used in most research assessments [26.9].

More specific to the context of altmetrics, an initiative to develop standards in altmetrics began in 2013, resulting in the *National Information Standards Organization (NISO) recommended practice, altmetrics definitions and use cases* [26.11]. The report comprises a detailed set of use cases in which the possibilities and limitations of a variety of altmetrics indicators for particular purposes by specific stakeholders is discussed. The NISO report also includes a code of conduct with respect to the responsible use of altmetrics data which focuses on transparency, replicability and accuracy of indicators.

26.3 Social Media Data and Indicators

The emergence of metrics of scholarly objects based on data from online social media platforms opened the possibility of analyzing new forms of interactions between different audiences and scholars (or scholarly products). These interactions are possible through the technical affordances allowed by these social media platforms, and have been conceived as "traces of the computerization of the research process" [26.21], resulting in the availability of different indicators based on user activity across the various online platforms. The NISO recommended practice, altmetrics definitions and use cases [26.11], defined altmetrics as:

> online events derived from activity and engagement between diverse stakeholders and scholarly outputs in the research ecosystem.

Social media metrics have also been discussed as a potential source of evidence in research evaluation, particularly in response to the quest for better metrics for measuring research performance [26.14].

Several advantages of social media metrics have been discussed, particularly over the more traditional approaches of research evaluation [26.10]. Among these advantages, *speed*, *openness* and *diversity* have been highlighted as some of the most important [26.10]. However, *Wouters* and *Costas* [26.10] also argued that for these new indicators to be realistically used in research evaluation, *transparency* and *consistency* are more important characteristics.

A theoretical framework for the use of altmetrics in evaluation was introduced by *Haustein* et al. [26.2]. Based on this framework, social media metrics can also be seen as:

> events on social and mainstream media platforms related to scholarly content or scholars, which can be easily harvested (i. e., through APIs), and are not the same as the more *traditional* concept of citations. [26.2]

This framework categorizes online acts upon *research objects*, including all forms of scholarly outputs (e. g., publications, but also data, code, etc.) as well as scholarly *agents* (e. g., scholars, funding agencies, research organizations). Thus, the realm of these new metrics would not be limited to the interactions with research outputs, but would include interactions with (and among) different scholarly agents, and the different forms of interactions could be characterized by the degree of engagement between the users with the scholarly objects.

However, in spite of these more conceptual discussions on the nature and characteristics of social media metrics, their strongly heterogeneous and ever-changing nature [26.22] has made the development of robust theories for the interpretation and evaluation of the activities and interactions captured by them very challenging.

26.3.1 Social Media Metrics Tools

In this section, the main characteristics of tools based on social media metrics are described. The purpose is not to discuss these tools as evaluative devices, but rather as sources of information on the relationships and interactions between science and social media. Thus, we take the approach that social media metrics are relevant sources for studying the interactions and relationships between science and social media, aligning more with what could be termed the *social media studies of science* [26.23] than with sources of scientific recognition or scientific impact. Moreover, our aim is not to focus on the currently available *altmetrics sources* but on the concepts behind these sources. Thus, although the current tools, sources and platforms collecting and providing social media data may disappear or change in the future (in what *Haustein* [26.22] has labeled the *dependencies* of altmetrics), many of the events and acts currently captured by *altmetrics data aggregators* could still be relevant in the future. For example, if Mendeley disappears, the idea of an online reference manager would still be feasible—with users from all over the world saving their documents—and counts of the number of different users (and types of users) saving these documents would still be possible should other new platforms be created. Moreover, while most common social media metrics tools usually refer to online events that exist around *scholarly outputs* (usually journal articles), there are also tools that focus on the activities of *scholarly agents*, particularly individuals. These tools and their main conceptual social media significance are described below:

- *Online reference management, social bookmarking and tagging tools*. Several online reference managers allow the counting of the number of times publications have been saved, bookmarked or tagged by different users of the platform. For instance, the *readership* counts provided by Mendeley (http://www.mendeley.com) include total number of users who have saved (added) a document to their private libraries. In addition, Mendeley offers some statistics on the academic status (students, professors, researchers, librarians, professionals, etc.), discipline and country of the users, as well as tags assigned to the publications they have saved. Other tools such as BibSonomy (https://www.bibsonomy.org/), Zotero (https://www.zotero.org) and CiteULike (http://www.citeulike.org/) also offer information on the posted counts/users, tags, posting history and user's info, plus the bibliographic information of the bookmarked or saved documents, although their APIs (application programming interfaces) are not yet fully developed [26.22].
- *Microblogging tools* (Twitter (https://twitter.com), Weibo (https://www.weibo.com), etc.) offer the possibility of disseminating information in small messages (e. g., the current 280-character limit for Twitter; before 2017 it was 140). In addition, these tools are aimed at broadcasting, filtering and establishing interactions among their users. For example, through the use of symbols such as @ or # in Twitter, it is possible to target other Twitter users (tweeters) and create messages (tweets) that are easy to filter or disseminate (retweet) to other users through the use of specific tags (the # symbol for thematic tags or the @ symbol to target other users). These tools also offer possibilities for *following* other users and *liking* (or appraising) other users' messages within the platform. Most microblogging tools offer the possibility of linking to external objects, which may be publications (e. g., through their Digital Object Identifier (DOI)) or other scholarly agents (e. g., scholars' websites, university websites). These technical options, or *affordances*, open the possibility to generate multiple indicators (e. g., the number of (re)tweets, likes, or followers around any particular scholarly object). An advantage of these platforms is that they provide rich information on users, tweets and locations through both their web interfaces and their APIs (Twitter streaming API, REST API with rate limit, or the commercial GNIP API (https://dev.twitter.com/docs) or Weibo open API (http://open.weibo.com/wiki/API%E6%96%87%E6%A1%A3/en)), thus making their data accessible and analyzable (although the different platforms may impose restrictions on the amount of data obtained).
- *Blogs and blog aggregators*. A number of blog platforms and blogging tools focus on peer-reviewed research, for example http://ResearchBlogging.org or http://ScienceSeeker.org. Blogs, and particularly scientific blogs, are an emerging means of disseminating discussions on scholarly materials [26.24] to other academics or the general public. Typical metrics that can be obtained from these platforms include blog mentions (e. g., the mentioning of a researcher or a university) or blog citations (e. g., citations to other scientific outputs). Information from blogging activities is usually available through their web interfaces or APIs.
- *Social recommendation, rating and review services*. Here we find some scholarly oriented tools such as F1000Prime (http://f1000.com/prime/about/whatis/how), which is a post-publication peer review service offering access to metrics such as views and downloads, as well as recommendation scores of biomedical literature, reviewed by their appointed users, together with information (labels or tags) on their type of recommendation (e. g., for teaching, controversial, new findings). Other academic platforms include Publons (https://publons.com/home/), which was recently acquired by Clarivate Analytics, and PubPeer (https://pubpeer.com/), which offer post-publication peer comments and scores for scholarly biomedical or multidisciplinary publications. A more general platform is Reddit (https://www.reddit.com/dev/api), which provides information such as comments and votes to the posts provided by its users. Some of these tools offer open APIs (Reddit), while for others (Publons or PubPeer) access is available only on request.
- *Wikis and collaborative content creation*. These platforms are seen as "collaborative authoring tool[s] for sharing and editing documents by users" [26.25]. A common metric available through these sources includes mentions of scholarly objects. For example, Wikipedia citations or mentions are available via its API (https://www.mediawiki.org/wiki/API:Main_page), enabling the analysis of the number of citations that scholarly publications have received in Wikipedia.
- *Social networking platforms* (e. g., LinkedIn (https://www.linkedin.com/), Facebook (https://www.facebook.com/)). These generalist platforms allow their users to connect, interact and communicate in many different ways (messaging, sharing, commenting, liking, etc.). Information on their users, activities and their geolocations are

typically available through their web interfaces or APIs (e. g., Facebook Graph and Public Feed APIs (https://developers.facebook.com/docs/graph-api) or LinkedIn API (https://developer.linkedin.com/docs/fields).

- *Social networking platforms for researchers* (e. g., ResearchGate (https://www.researchgate.net/) and Academia.edu). These tools provide information on scholars and their outputs and affiliations, and offer different metrics at the individual, institutional or country levels. This type of platform, inspired by the more generalist social networking platforms, aims at facilitating networking and communication among scholars, finding academic content, experts or institutions, and as sharing and disseminating their research with peers. ResearchGate (RG) offers different indicators including the RG Score (a measure of reception of a researcher's publications and her participation on the platform) and RG Reach (a measure of visibility of a researcher's publications on the platform), together with other indicators such as the number of citations, reads, downloads, h-index and profile views. It seems that the RG Score is influenced by a researcher's academic and online activities, and hence it is suggested to reflect a combination of scholarly and social networking norms [26.26, 27]. Other platforms such as Academia.edu provide information on mentions of a researcher's name by others, on the *readers* (including views, downloads and bookmarks of a researcher's publications), profile views and visitors per date, country, cities, universities, job titles, etc., some of which are available by monthly subscription.
- *Altmetrics data aggregators*. These are tools such as Altmetric.com, Lagotto (http://www.lagotto.io/), PLOS ALM (Article-Level Metrics) (https://www.plos.org/article-level-metrics), Plum Analytics (http://plumanalytics.com/) and ImpactStory (https://impactstory.org/) which aggregate metrics for scholarly materials from different sources. Examples of the metrics provided by these aggregators include *views*, *saves*, *citations*, *recommendations* and *discussions* around scientific publications by PLOS ALM and Lagotto, or those of *usage*, *captures*, *mentions*, *social media* and *citations* by Plum Analytics. Altmetric.com provides a composite weighted indicator (*Altmetric Attention Score*) of all the scores collected around scientific outputs (https://www.altmetric.com/about-our-data/the-donut-and-score/). Although most of these aggregators are based on a similar philosophy (to capture online events around scholarly objects), they often differ in the sources they track (publications with a DOI or PMID [PubMed identifier], etc.), the methodologies they use to collect the data (using public or commercial APIs, etc.) and the way they process and report the metrics (e. g., raw vs. more aggregated indicators). They usually also differ in terms of their updates, coverage and accessibility [26.28].

26.3.2 Characterizing Interactions and Users in Social Media Metrics

The relationships between scholarly objects and social media users can be characterized from two different perspectives: the *typologies of social media users* who interact with the scholarly objects, and the *typologies of social media interactions* that are established between the social media users and the scholarly objects:

- *Typologies of social media users*. The analysis of social media users has been approached from various perspectives, and a general framework (unified media-user typology) has been suggested for unifying all media user types based on user frequency, variety of use and their content preference [26.29]. According to [26.29], the term *user typology* is defined as the:

  > categorization of users into distinct user types that describes the various ways in which individuals use different media, reflecting a varying amount of activity/content preferences, frequency and variety of use

  which could be influenced by psychological, social and cross-cultural factors [26.29, 30].
  In the realm of social media metrics, different user typologies have been identified in the literature. For example, Mendeley users have been studied based on the information that they have provided about themselves on Mendeley (self-classified as *students*, *researchers*, *professors*, etc.) [26.31–33]. Tweeters have also been categorized as *influencers/brokers*, *discussers/orators*, *disseminators/bumblers* or *broadcasters*, based on the combination of the number of followers and their engagement with the publications [26.34–36]. Altmetric.com also categorizes tweeters as *researchers*, *science communicators*, *practitioners* or *general public*, based on the tweeters' descriptions. Other efforts have focused on the study of scholars active on Twitter [26.37–39].
- *Typologies of social media interactions*. How social media users interact with the scholarly objects can provide valuable information with which to characterize the indicators. *boyd* and *Ellison* [26.40]

argued that although social media tools have some common features (such as creating a profile for making connections), they differ in terms of the way users interact with the platform. For example, *bridging* and *bonding* refer to different forms of ties established among different users on social media [26.41, 42], based on the following/followees model in Twitter [26.43]. Thus, according to *Hofer* and *Aubert* [26.42], the use of Twitter is mainly influenced by *bridging* ties (i. e., following users from different networks with the aim of broadening the information flow) rather than *bonding* (i. e., following like-minded people for gaining emotional support). This form of follower/followee interactions is also very central in several science-focused altmetrics platforms such as ResearchGate or Mendeley. *Robinson-Garcia* et al. [26.44] have proposed the analysis of the relationship of follower/followees on Twitter as a means to identify potential traces of societal interactions. Another example includes the analysis of interactions via other social media platforms (like Facebook) between students and their instructors [26.45]. More focused on the context of social media metrics, *Haustein* et al. [26.46] established three main categories of engagement (or interaction) between the users and the scholarly objects: *access* (related to viewing, downloading and saving), *appraise* (mentioning, rating, discussing, commenting or reviewing) and *apply* (using, adapting or modifying). Typologies of blog posts have been discussed based on the content and motivations of the bloggers (e. g., discussions, criticisms, advice, controversy, triggers) [26.47]

26.4 Conceptualizing Social Media Metrics for Research Evaluation and Management

In order to discuss potential uses of social media metrics, we need to understand the reliability and validity of social media indicators for evaluative purposes. Section 26.4.1 discusses the criteria that social media indicators should meet in order to be considered valid indicators. Section 26.4.2 explains to what extent indicators should be homogeneous in their composition [26.48]. Finally, the dependence of social media metrics on external data providers and the technical quality of the data is discussed in Sect. 26.5.

26.4.1 Validity and Reliability of Social Media Metrics

In the discussion around the possibility of altmetrics as new sources of indicators for research evaluation, *Wouters* and *Costas* [26.10] suggested that altmetrics "need to adhere to a far stricter protocol of data quality and indicator reliability and validity". According to *Gingras* [26.48], in order to be valid, indicators should meet three essential criteria:

1. *Adequacy*
2. *Sensitivity*
3. *Homogeneity*.

The concept of validity relates to an indicator's success in measuring what is expected to be measured [26.49]. The notion of adequacy indicates how the indicator captures the reality behind the concept intended to be measured. Along similar lines, as suggested by *Nederhof* [26.50] regarding bibliometric indicators, the main question is to what extent social media indicators are valid as measures of research performance. In scientometrics, citations have been assumed to be imperfect proxies of intellectual influence or scientific impact. This imperfection is derived from the fact that quite often this is not the case, citations may be perfunctory, and the choice of citations involves a substantial degree of arbitrariness by the authors, thus deviating from the idea of citations as measures of intellectual influence [26.51–54].

In the case of social media metrics, this issue is more complicated, as it is not clear to what extent these indicators are even remotely related to the concept of scientific impact. On the one hand, indicators such as Mendeley readers or F1000Prime recommendations have a closer relationship with scientific impact, as they have a strong scholarly focus. Indicators derived from platforms such as ResearchGate or Academia.edu can also be expected to have a closer conceptual link to the traditional concepts of scholarly impact and performance. However, the lack of studies based on these platforms renders any consideration of them merely tentative. On the other hand, social media indicators derived from sources such as Twitter or Facebook are more difficult to relate to the concepts of scientific impact and scholarly activities. These indicators are usually thought of as measuring types of interactions that are not (directly) related to research performance.

The second criterion highlighted by *Gingras* [26.48] is sensitivity or inertia, understood as the *resis-*

tance to change of indicators. According to this notion, a good indicator should vary "in a manner consistent with the inertia of the object being measured". In the case of traditional bibliometric indicators, they usually have a slow inertia. They typically don't suffer from sudden and drastic changes, and although there are sources that may distort some of the indicators, most of them respond to an inertia that seems to align with the common perceptions of how scientific impact or performance also changes. Mendeley readership and F1000Prime recommendations have a similar inertia as citations [26.55–57]. However, the sensitivity and inertia of social media metrics can be challenged by three main issues:

- *Speed*. Traditionally considered one of the most important advantages of social media metrics, as they tend to happen faster than citations, their speed is also one of their most important limitations [26.10]. For example, indicators based on social media platforms like Twitter can change dramatically in a matter of hours as a result of controversies triggered by the publications, mistakes in the papers, or even jokes.
- *Superficiality*. The faster nature of most social media metrics may indicate a lower engagement of the users with the scholarly objects, which may be related to a higher level of superficiality in the appraisal of the objects. For example, Twitter users may massively (and suddenly) (re)tweet a publication without any intellectual engagement with it.
- *Small changes*. Given the fact that many of these indicators tend to present low values [26.46], small changes in the values of the indicators could have large effects. For example, a small increase in the number of (re)tweets, or a few additional mentions in blogs, may cause substantial changes in the indicators (e. g., drastically increasing their percentile value). Due to the strong skewness of most social media indicators [26.58], for most publications, just a few additional scores would propel a publication from a lower percentile to a higher percentile. For example, the paper https://www.altmetric.com/details/891951#score was tweeted by just two Twitter users on 15 December 2017, which positioned the paper in the 54th percentile according to Altmetric.com, while the paper https://www.altmetric.com/details/3793570#score was mentioned by four tweeters (i. e., just two additional tweeters), classifying it in the top 25th percentile (on 15 December 2017). These examples illustrate the strong sensitivity of these indicators to small changes, also illustrating the ease with which they can be manipulated [26.10, 59].
- *Reliability*. The sensitivity notion described by *Gingras* [26.48] can also be related to the reliability of indicators. Reliability is the extent to which an indicator yields the same result in repeated measurements. In the case of bibliometrics, the citation process is considered to be stochastic [26.50]. Papers of equal impact do not necessarily receive identical numbers of citations, since multiple random factors come into play (e. g., biases of the citers, publication and citation delays, coverage issues). Social media metrics are generally less reliable due to the stronger dependence on the consistency and accuracy of the data collection methodologies [26.28] and the low coverage of publications by social media sources [26.46, 60].

26.4.2 Homogeneity (or Heterogeneity) of Altmetric Indicators

This concept of homogeneity is especially important with respect to composite indicators that combine different measurements into a single number, thus "transforming a multidimensional space into a zero-dimension point" [26.61], although composite indicators are still possible when important mathematical and conceptual limitations are met. Research has shown significant heterogeneity in social media metrics [26.2, 10, 22] and a variety of relationships among them [26.34, 35]. In general, citations and Mendeley readerships are the most closely related indicators [26.62, 63]. Similarly, F1000Prime reviews are conceptually similar to peer review indicators [26.64, 65]. However, indicators based on Twitter, blogs or news media are both conceptually and empirically different from citations [26.60, 66] and also differ among themselves. These indicators capture different types of impacts. Therefore, constructing composite indicators and mixing these indicators for research evaluation should be discouraged. Maintaining the various altmetrics scores as separate entities is the best choice for ensuring transparency in assessment approaches. Examples of composite altmetrics indicators include the Altmetric Attention Score and the RG Score, which lump together fundamentally different metrics (Twitter, blogs, views, etc.) [26.2]. Although the calculation formula for the Altmetric Attention Score is disclosed (unlike the RG Score, which has remained a black box), the validity and application of this composite indicator for evaluative purposes is unclear.

In addition, we would like to call attention to problems related to the lack of *internal homogeneity* within a single indicator for many social media indicators. Perhaps the clearest example is the inclusion of tweets and retweets in the same indicator. Although both tweets and retweets come from the same platform, they arguably have different roles and should therefore be valued differently [26.67]. Other examples include the count for all of Mendeley readership in the same indicator, combining academic users (professors, PhDs, etc.) with nonacademic ones (e. g., librarians, professionals, students), or the aggregation of Facebook shares, likes and comments in a single indicator [26.22]. A lack of internal homogeneity may have a dramatic effect on the comparison of metrics from different data aggregators [26.28]. Therefore, transparency on the part of data providers in how indicators are structured and calculated is fundamental to the ability to judge the validity and replicability of social media metrics [26.22].

26.5 Data Issues and Dependencies of Social Media Metrics

As pointed out by *Haustein* [26.22], a central issue that must be considered for any application based on social media metrics is the direct dependence on altmetrics data aggregators, which themselves are dependent on other major social media data providers (Twitter, Facebook, etc.). Thus, any application of social media metrics is potentially limited by the decisions, strategies and changes on the part of any of these actors [26.68]. As a result, variations in their policies may mean the disappearance of a data source (e. g., in recent years of the existence of Altmetric.com, sources such as Sina Weibo and LinkedIn have stopped being covered, and the online reference manager Connotea has been discontinued [26.22]), restrictions on a type of analysis (e. g., current data restrictions regarding dates in Mendeley hampers analysis of readership trends) or a complete modification of the concept of impact or activity being measured (e. g., the conflating of posts, shares and likes from Facebook in a single indicator may confound the meaning of the indicator). Regarding data quality issues, a critical limitation is the dependence on unique identifiers of scientific publications such as DOI or PMID. Publications without any of these identifiers are excluded from the tracking algorithms of altmetrics data aggregators. Mentions of scientific publications must also include a direct link to the scientific publication. Mentions of publications using only their titles or other textual characteristics, or links to versions of the publication not covered by the altmetrics data aggregators, will be ignored.

26.6 Conceptualizing Applications of Social Media Metrics for Research Evaluation and Management

In this section we conceptualize some applications of social media metrics. Although most of our examples are taken from actual practices, the aim is to provide a perspective that could transcend current tools and databases. Thus, regardless of the future availability of the current tools, we consider that most conclusions would remain relevant, should similar tools (or variations of current tools) still be in place and accessible.

In order to provide a comprehensive conceptualization of applications of social media metrics, we need to discuss the main types of possible applications. In the field of bibliometrics, a differentiation has been made between *descriptive bibliometrics* and *evaluative bibliometrics* [26.69–71]. According to *Van Leeuwen* [26.71], descriptive bibliometrics are related to top-down approaches able to provide the *big picture*. This more descriptive notion of bibliometrics is also related to the contextual perspectives recently proposed in scientometrics [26.72]. We speak of evaluative bibliometrics if bibliometrics is used to assess the research performance of a unit of analysis (research teams, research organizations, etc.), often in a comparative framework. For example, different units can be compared in terms of citations or publications, or a unit can be compared with a specific benchmark (e. g., the average citation impact in the field(s), as is done for *field-normalized* indicators). The problem with the descriptive/evaluative dichotomy is that it is not always possible to clearly distinguish the two approaches. In practical terms, any bibliometric description can become an evaluative instrument. For example, the mere reporting of the number of publications of a university department may turn into an evaluative indicator if is compared to other departments (or a benchmark) and used, for example, to allocate resources.

Therefore, we propose that a distinction be made between *descriptive* and *comparative* approaches. As descriptive approaches, we consider those approaches that focus on the analysis and description of the activities, production and reception of scholarly objects for different units of analysis, together with the analysis of the dynamics and interactions among different actors and objects. As comparative approaches we consider those approaches that are (mainly) focused on the comparison of outputs, impacts and actors, often in the context of evaluation. Simply put, descriptive approaches are related to questions of *who*, *when*, *how* and *what*, while comparative approaches are concerned with questions of *fast(er)/slow(er)*, *high(er)/low(er)*, *strong(er)/weak(er)* or just *better/worse*. Of course, comparative approaches are by definition based on some form of descriptive input data. Both descriptive and comparative approaches can be used as tools in research evaluation, but they can also be used for other purposes (e. g., knowledge discovery).

Social media metrics have typically been discussed in light of their potential role as a replacement for citations for comparative and evaluative purposes [26.1]. Less research has focused on the potential value of social media metrics from a more descriptive perspective. In Table 26.1 we summarize a general framework of potential applications for social media metrics based on the descriptive/comparative dichotomy.

26.6.1 Descriptive Social Media Metrics

As shown in Table 26.1, descriptive approaches use basic analytical indicators such as total count summaries, trend analysis and thematic landscapes, as well as network approaches that consider the dynamics and interactions between different social media agents and scientific outputs. Similar to bibliometric indicators, descriptive indicators can be calculated with the objective of identifying general patterns in social media reception of scientific publications of a given unit. In Table 26.2 we present an example: basic descriptive indicators for three major data sets comprising publications from Africa, the European Union (EU28) and the United States (USA) covered in the Web of Science (WoS) for the period 2012–2014, and that have a DOI or PMID.

We would like to emphasize that certain elements must be taken into account when reporting social media metrics. It is important to disclose the total output analyzed (indicator *P* in Table 26.2). In our case, as we have worked with data collected from Altmetric.com (until June 2016), only publications with a DOI or a PMID have been tracked in this source. Thus, the data set is reduced to only publications with an identifier traceable by this data provider (indicator *P*(DOI/PMID) in Table 26.2).

In the second section of the table, we explore the total social media counts that are obtained for each of the sets of publications. Thus, TTS counts all the Twitter mentions (in this case combining both original tweets and retweets) for the publications. TBS is the total blog citation score, TNS is the total news media mentions score, TPDS is total policy document citations score and TWS is the total Wikipedia citations score. There are other indicators that also could have been calculated based on Altmetric.com, such as those based on Facebook, Google Plus or F1000Prime. For a discussion of other social media metrics, we refer here to *Costas* et al. [26.73].

In the third part of the table, we calculate the averages of the different scores per publication. Simply put, each of the total scores is divided by the number of publications that could be tracked (*P*(DOI/PMID)). Thus, we can talk about the mean Twitter score (MTS), for example, or the mean blog score (MBS). Obviously, the mean is not necessarily the only statistic we could have calculated. Other descriptive statistics such as median, mode or min–max values could have been obtained.

Finally, in the fourth section of the table, we present another possibility for basic social media metrics. Given the strong skewness of most altmetrics indicators [26.58, 74] as well as their sparsity [26.75], mean values can be strongly influenced by outliers (e. g., extremely highly tweeted publications), an issue that is not uncommon among this type of indicator [26.60]. In addition to the use of median- or percentile-based indicators to help mitigate the problem, indicators of the coverage of the publications with a given degree

Table 26.1 Conceptualization of descriptive and comparative *social media metrics* approaches

Descriptive social media metrics	Comparative social media metrics
• Descriptive social media indicators (nonnormalized), e. g.: – Total counts, coverage – Trend analyses • Social media metrics landscapes – Thematic landscapes – Geographic landscapes • Network approaches, e. g., communities of attention, Twitter coupling, hashtag coupling	• Normalized indicators, e. g.: – Mendeley field-normalized indicators – Percentile-based indicators (e. g., Altmetric Attention Score) • Social media-based *factors* (e. g., Twimpact factor, T-factor) • Composite social media indicators (e. g., RG Score, Altmetric Attention Score) • Comparative network indicators (e. g., relative centrality)

Table 26.2 Examples of basic descriptive altmetrics indicators for Web of Science publications (with a DOI or PMID) from Africa, EU28 and USA (2012–2014)

1) Output					
Unit	***P***	***P* (DOI/PMID)**			
Africa	125 764	104 008			
EU28	1 605 393	1 305 391			
USA	1 686 014	1 281 624			
2) Total counts					
Unit	**TTS**	**TBS**	**TNS**	**TPDS**	**TWS**
Africa	190 737	6126	11 291	886	2154
EU28	2 034 833	67 262	118 568	4153	23 126
USA	3 461 227	136 682	263 517	4964	32 647
3) Averages					
Unit	**MTS**	**MBS**	**MNS**	**MPDS**	**MWS**
Africa	1.83	0.06	0.11	0.01	0.02
EU28	1.56	0.05	0.09	0.00	0.02
USA	2.70	0.11	0.21	0.00	0.03
4) Coverage					
Unit	**PP(t1) (%)**	**PP(b1) (%)**	**PP(n1) (%)**	**PP(pd1) (%)**	**PP(w1) (%)**
Africa	27.0	2.7	2.1	0.6	1.2
EU28	28.5	2.7	2.3	0.2	1.2
USA	37.4	5.1	4.5	0.3	1.8

P: Total publications of the unit
P(DOI/PMID): No. of publications with a DOI or a PubMed ID
TTS: Total Twitter mentions score
TBS: Total blog citations score
TNS: Total news media mentions score
TPDS: Total policy document citations score
TWS: Total Wikipedia citations score
MTS: Mean Twitter mentions score
MBS: Mean blog citations score
MNS: Mean news media mentions score
MPDS: Mean policy document citations score
MWS: Mean Wikipedia citations score
PP(t1): Proportion of publications with at least one tweet mention
PP(b1): Proportion of publications with at least one blog citation
PP(n1): Proportion of publications with at least one news media mention
PP(pd1): Proportion of publications with at least one policy document citation
PP(w1): Proportion of publications with at least one Wikipedia citation

of metrics can be provided. In Table 26.2 we give the proportion of publications that have at least *one* mention in each of the metrics (i. e., one tweet, one blog citation, etc.). Thus, we can see that about 27% of African publications (with a DOI/PMID) have been tweeted at least once, while 5.1% of all US publications (with a DOI/PMID) have been cited at least once in blogs. The use of the at-least-*one*-mention option (represented by the value 1) coincides with the absolute coverage of publications in each of the social media sources. However, this value of 1 could have easily been changed by any other value (e. g., 2, 3, a particular percentile, the number of only original tweets (i. e., excluding retweets)). Moreover, coverage indicators can also be subjected to normalization (e. g., the equalized mean-based normalized proportion cited (EMNPC), as suggested by [26.75]); however, more complex indicators such as these introduce a more comparative nature, in which the coverage of units is compared to a global reference.

Trend Altmetrics Indicators

In addition to the basic indicators discussed above, it is possible to provide trend analysis (Fig. 26.2), giving social media time series data with properties that differ from bibliometric indicators. However, the data collected by most of the altmetrics data aggregators are very recent, and the application of trend analysis is therefore relatively limited. Moreover, uncertainties regarding methodological changes in social media data

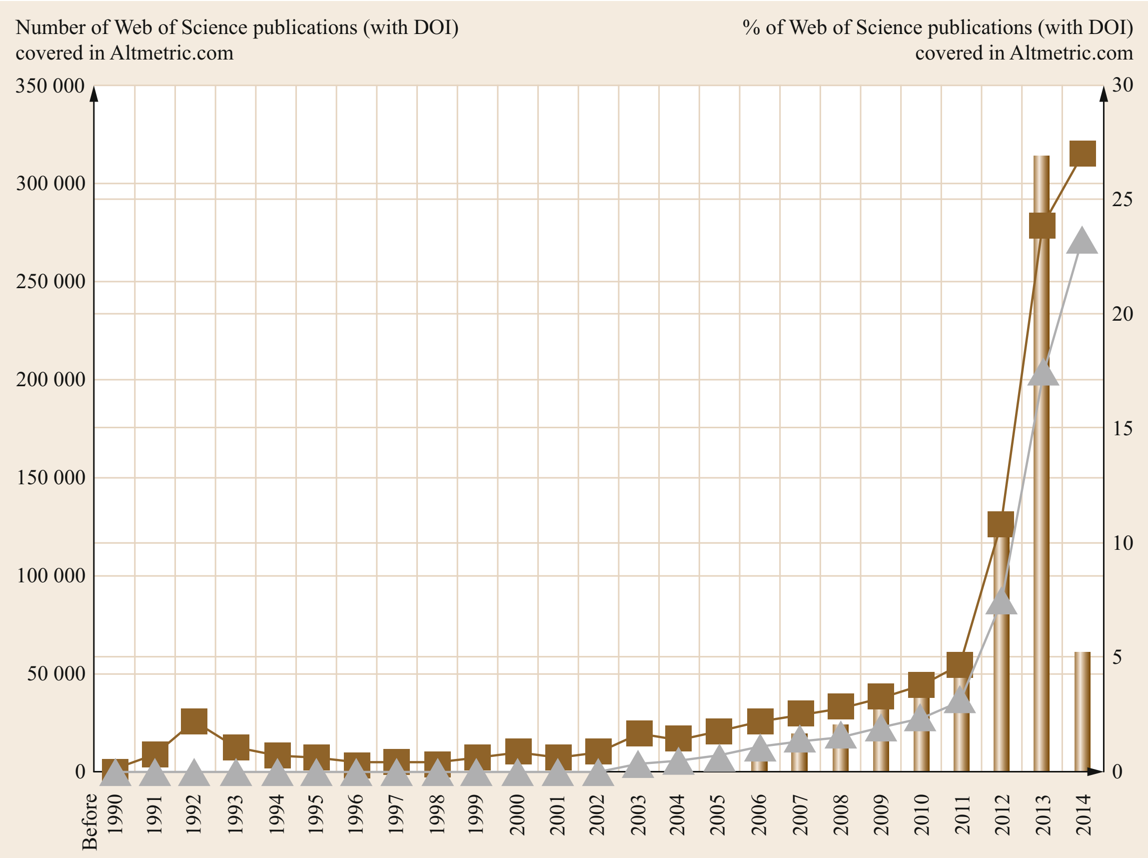


Fig. 26.2 Number and share of publications from Web of Science (DOI) with coverage in Altmetric.com, 1980–2013 (after [26.60], with permission from Wiley). Altmetric.com started their data collection in July 2011

collection warrant caution in the interpretation of trend analysis. For example, trend analyses may be influenced by improvements in the algorithms for identifying mentions of scientific publications by the altmetrics data aggregators, thus not reflecting genuine trends in the indicators themselves.

Although Mendeley data are conceptually close, albeit not identical, to citations, their time series properties are very different [26.55–57]. This can be seen in Fig. 26.3 below. In contrast to citations, which are generally higher (and never decrease) as time goes by, Mendeley readership values can decrease, as Mendeley users can delete publications from their libraries or fully erase their Mendeley profiles.

Longitudinal Analysis—Social Media Histories

Similar to citation analysis, in which it is possible to study the impact of scientific publications longitudinally over time (in so-called *citation histories* [26.76]), *social media* or *reception histories* are also possible. Examples are the analysis of the accumulation of Mendeley readership, blog citations or tweets over time for any set of publications. The time stamps of the tracked events are generally highly accurate (e. g., the exact time a tweet was sent, or when someone saved a document in her Mendeley library), thus enabling longitudinal trend analysis. However, the following problems challenge the development of longitudinal analysis of social media metrics:

- *The lack of openly available diachronic information*. In the case of Mendeley, concrete information on when the readership was produced is not available through their public API. This creates difficulties in both the calculation of longitudinal readership analysis and the potential determination of *readership windows* (e. g., variable or fixed windows could also be established, similar to citation windows [26.77]). This lack of diachronic information about Mendeley readership hinders the development of studies on the potential predictive power of early Mendeley readership for later cita-

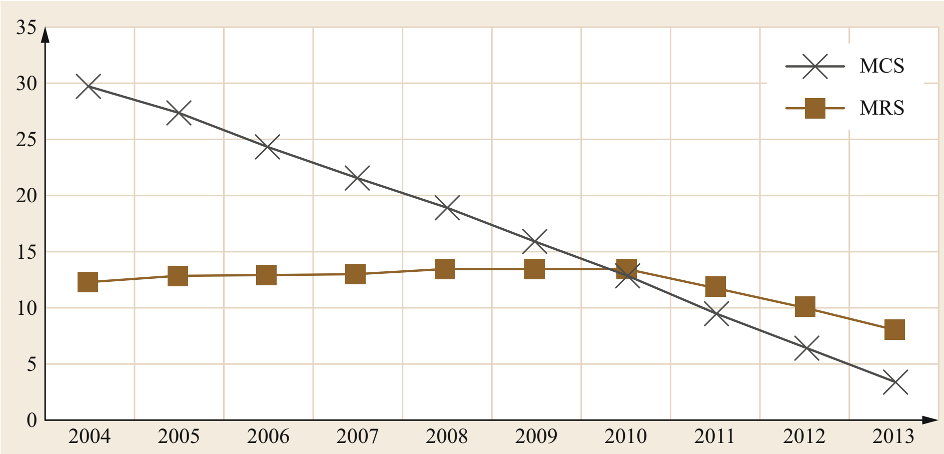


Fig. 26.3 Distributions of mean readership score (MRS) and mean citation score (MCS) indicators for the WoS publications over time (x axis shows the publication years and y axis shows the mean scores for citations and readership) (after [26.57], with permission from Wiley)

tions. A possible solution is the repeated tracking of readership counts for publications over time, as was done for example in [26.56, 78].

- *Indeterminate publication time of scientific outputs*. Although in bibliometrics the use of the publication year of scientific outputs is the most common approach for determining the starting moment of a publication, there are important inconsistencies in the publication dates of scientific articles [26.35]. This is caused by the gaps between the actual moment a publication becomes accessible to the public (e. g., through the *online first* option of many publishers, or through its publication in a repository) and the official publication in a scientific venue (e. g., journal, conference, book). These inconsistencies are even more challenging when working with social media metrics. Given that social media interactions usually take place earlier and faster than citations, accurate knowledge of the actual time that a publication became available to the public is critical to establishing accurate time windows for the analysis of the social media reception of publications.

Social Media Metrics Landscapes

The possibility of providing different types of analytical landscapes based on social media metrics is one of the most interesting types of descriptive approaches. Conceptually speaking, there are two general landscape typologies: thematic landscapes and geographic landscapes (the two can be combined).

Thematic Landscapes. In scientometric research, thematic classification is an important asset, enabling the analysis of the structure and dynamics of scientific disciplines [26.79]. In media research, the introduction of thematic perspectives is also important. Social media metrics (e. g., Twitter, Facebook) have a stronger presence among social sciences and medical and health sciences [26.46, 73]. Figure 26.4 gives an example of an advanced social media thematic landscape. It presents tweets to all African and EU28 countries' publications (same publications as discussed in Table 26.2) using a publication-level classification comprising more than 4000 *micro-fields* and described in [26.79]. This is the same classification scheme used for the field-normalization of citation indicators applied in the Leiden Ranking (http://www.leidenranking.com/information/indicators). The size of the nodes represents the African and EU28 outputs published in that particular micro-field, while the color represents the share of those publications that have received at least one tweet (this is the indicator PP(tw1) discussed in Table 26.2). The nodes (fields) are positioned on the map according to their direct citation relations using the VOSviewer clustering method as described in [26.79, 80], based on the overall Web of Science database (period 2000–2016).

In Fig. 26.4, some of the most important topics of both African and EU28 research can be seen on the left-hand side of the map, which is the part of the map that concentrates most health-related and social sciences topics. The differences between public and scientific interest in topics between Europe and Africa become visible on these maps. Twitter reception of Africa's output gives priority to HIV-related topics as well as diseases such as tuberculosis or malaria. Other topics with a strong presence on Twitter with African participation refer to the ATLAS collaboration and the Higgs boson research (right-hand side of the map). In EU28 countries, psychological issues (emotions, depression, bulimia), cancer and obesity are among the main topics with large scientific production and strong presence on Twitter [26.81].

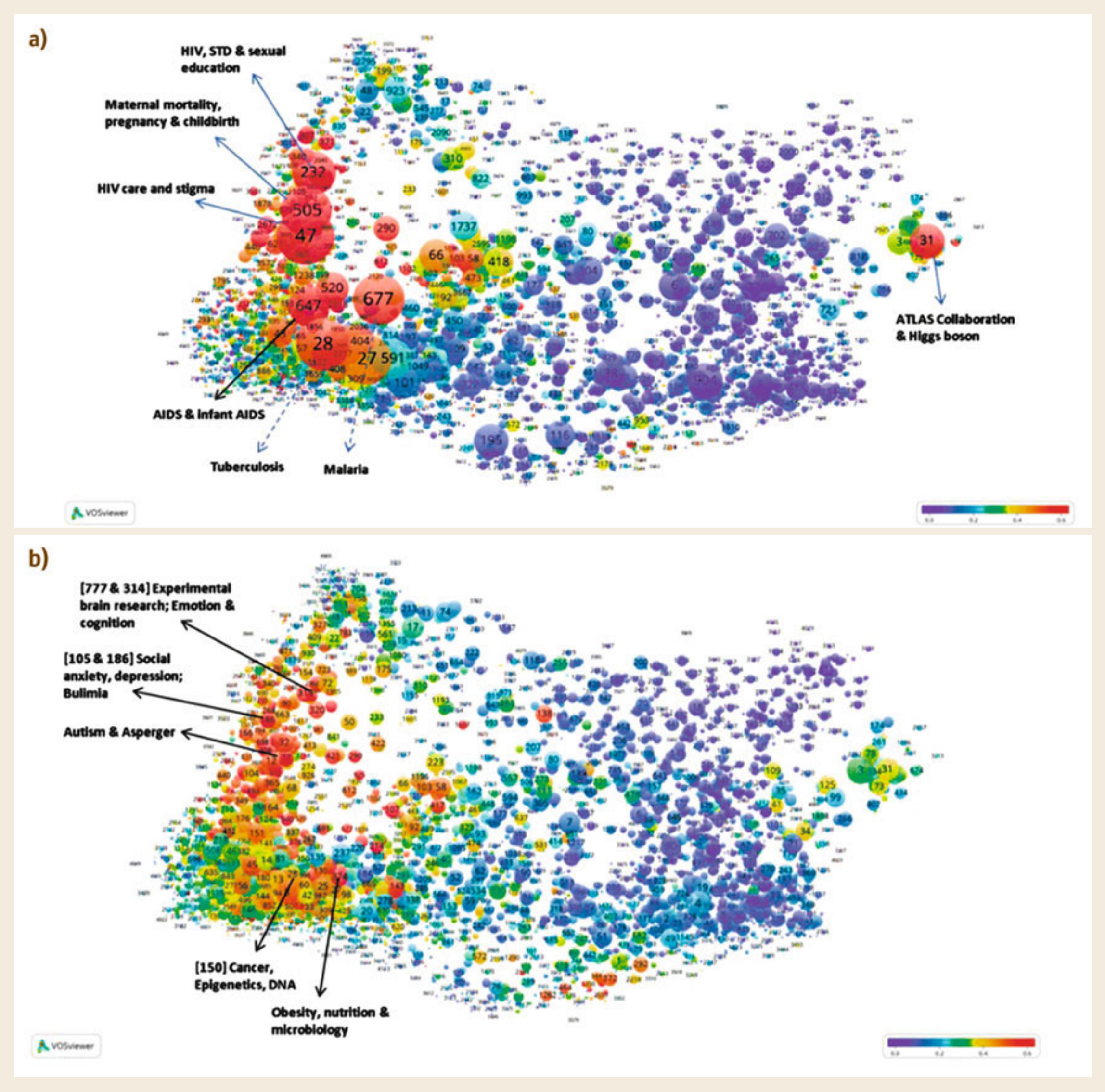


Fig. 26.4a,b Tweets thematic landscape of African publications **(a)** and EU28 publications **(b)**. Nodes represent fields (clusters of publications closely related by direct citation relations) and position on the map by the strength of their citation relations

Geographic Landscapes. In addition to thematic landscapes, it is possible to introduce a geographic dimension in the analysis of social media metrics. The geography can be determined by the geo-location of the entities reflected in the publications under analysis (e. g., authors, affiliations, funders, journals, or even the geography of the research itself, such as malaria in Africa researched by Dutch scholars). Alternatively, the geo-location of the different types of users who interact with the publications through the various social media platforms can serve as the basis for the landscapes. Thus, it is possible to study what the Mendeley users from South Africa read, or what publications are being tweeted from Nigeria. This particular type of analysis has two fundamental challenges: 1) the lack of disclosure of geographic information for all social media users (e. g., not all users in Mendeley, Facebook or Twitter disclose their geo-location), and 2) the variable granularity of available geographic information (e. g., not all users disclose their full geographical information; some provide only country-level information, while others also disclose region or location).

Figure 26.5 presents a world map showing the share of publications with at least one tweet (i. e., the PP(tw1) indicator as discussed in Table 26.2) across the countries of the authors. Red indicates higher PP(tw1) values, and blue indicates lower values.

As shown in Fig. 26.5, several African countries have a relatively high proportion of publications mentioned at least once on Twitter. Publications from Anglo-Saxon (e. g., USA, UK, Australia) and Northern European countries (e. g., the Netherlands, Denmark) are also tweeted frequently. The indicator PP(tw1) presented in Fig. 26.4 does not consider differences between fields, years or languages. Therefore, only the major patterns of the share of publications with some Twitter discussion can be extracted from it. However, the graph could also be obtained normalizing by fields, periods of time, or tweets from relevant tweeters (e. g., academic tweeters or tweeters from the same country as the authors of the papers).

Network-Based Indicators

The third type of descriptive social media metrics comprises network-based approaches. These are focused on analyzing the relationships and interactions among the different actors. These are the least developed, and more research will be necessary to fully grasp the possibilities of these analyses. In this section we will focus on just three basic examples of current applications:

1. The analysis of *communities of attention* [26.34]
2. *Hashtag coupling analysis* [26.82]
3. *Reading/reader pattern analysis* [26.83–86].

Communities of Attention. The analysis of *communities of attention* refers to the analysis of different communities of users active in social media platforms (e. g., tweeters, bloggers, Facebook users, etc.), and their interactions with scientific outputs or entities. This type of analysis goes beyond the analysis of *follower/followees* that many platforms allow, to include other types of interactions. Figure 26.6 presents the example of the Twitter community of attention for the set of African publications discussed in Table 26.2. In this network map tweeters are clustered together when they tweet the same publications, thus suggesting common scientific interests among them.

Figure 26.6 shows several clusters of Twitter users (communities) around African publications. Specifically, there is a strong user cluster (around @HIV_insight) with clear interest in HIV research, surrounded by other Twitter users related to AIDS research and sexual and medical topics. The yellow cluster

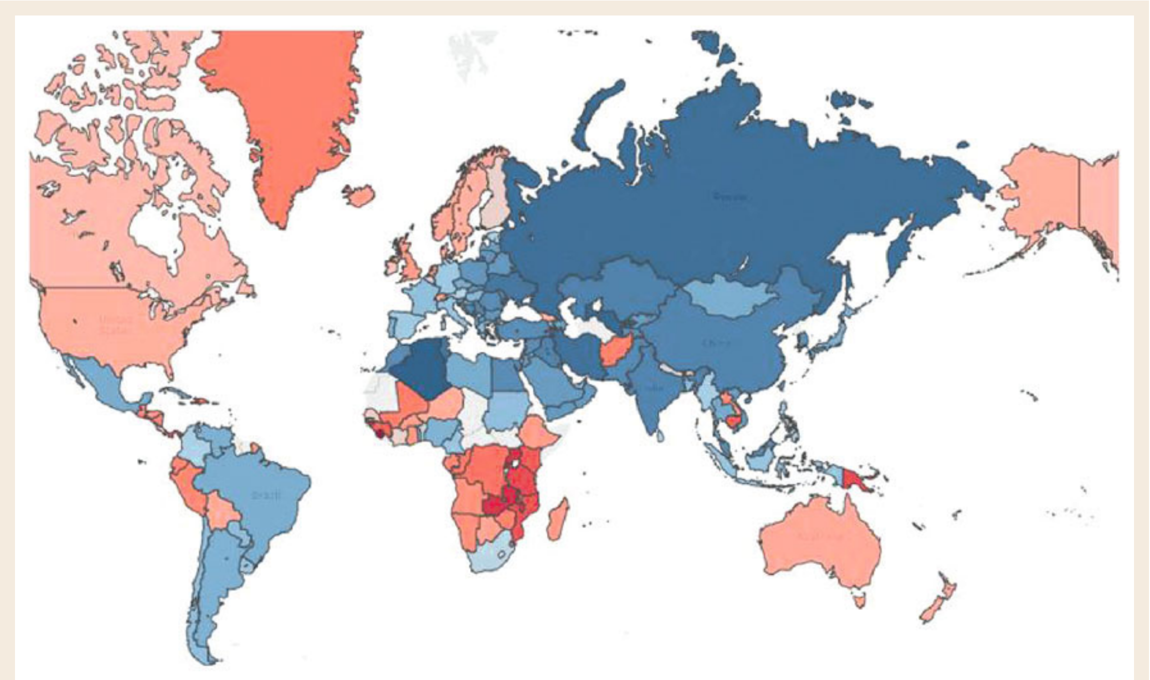


Fig. 26.5 Global map of the share of WoS publications (with a DOI/PMID, period 2012–2014) with at least one Twitter mention (PP(tw1)) across the countries of the authors. Threshold for *red*/*blue* differences is 34% (i. e., PP(tw1)< 34% *blue*, PP(tw1)≥ 34% *red*)

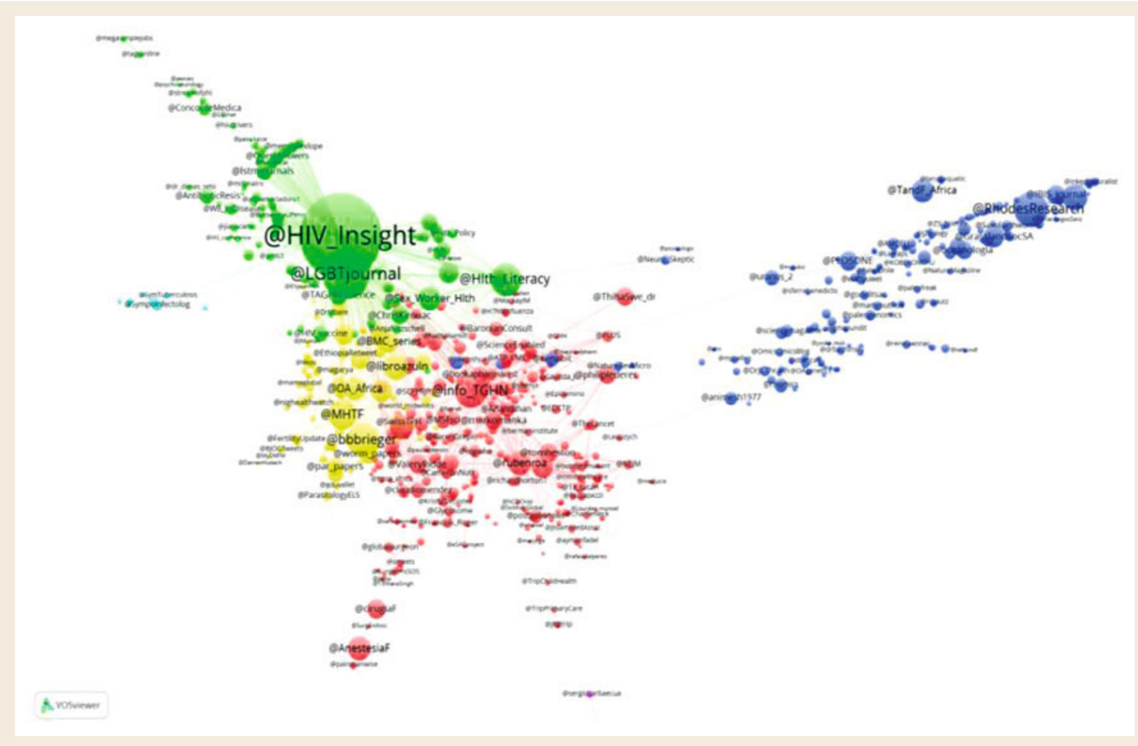


Fig. 26.6 Main Twitter community of attention map of African publications. Nodes are Twitter users; linkages/proximity of the nodes is determined by the number of common publications they have tweeted. Position of nodes in the map: VOSviewer clustering method

combines multiple users related to publishing issues. The dark blue cluster concentrates multiple users from a more multidisciplinary nature (e. g., the Twitter account of PLOS ONE). Conceptually speaking, this type of analysis does not need to be restricted to Twitter. It can be applied to any type of social media users (e. g., bloggers, Facebook users, Mendeley users).

Hashtag Coupling Analysis. This analysis is based on the *hashtag* affordance available on Twitter. Hashtags are used by Twitter users to link their tweets to broader *conversations*, expanding the potential exposure of their tweets to users beyond their original set of followers. When tweeters link the same set of publications to different hashtags, they are creating a network of related conversations. This type of analysis enables the study of the different existing conversations around scientific topics and can inform communication offices, students or researchers about specific hashtags related to their scientific topics or areas interest. It may also help scholars interested in disseminating important scientific results on Twitter to improve their communication strategy (e. g., by liking their tweets and publications to relevant hashtags). In Fig. 26.7, an example of Twitter hashtag coupling analysis is presented for the most frequent hashtags linked to scientific publications covered by Altmetric.com [26.82]. In the blue cluster it is possible to see how research linked to #prostatecancer or #oncology has also been linked to the broader hashtag #cancer. Similarly, #openaccess and #OA (green cluster) are coupled, as they are linked to a similar set of publications.

Reading/Reader Pattern Analysis. Data extracted from reference manager tools such as Mendeley or CiteULike have been used for knowledge domain detection or for finding common interests among their users [26.84, 87]. The idea is similar to cocitation [26.88, 89]. Those publications with high cooccurrence among different users' profiles are considered to be more similar in terms of their thematic subject [26.84]. The network of user groups in Mendeley saving the same set of publications showed that students and postdocs have more common topical interests than other user groups [26.83]. Others visualized readership activities and topics of interest among Mendeley users using the text mining functionality of VOSviewer, and

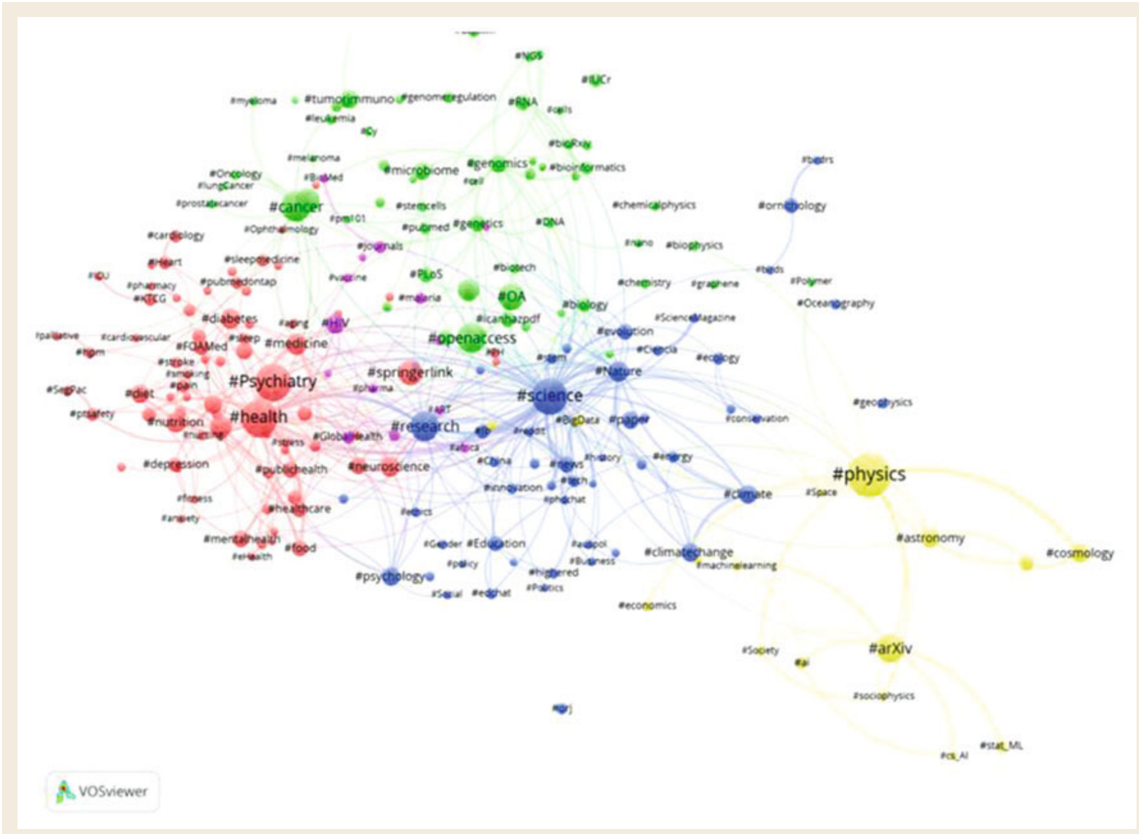


Fig. 26.7 Network map of the most common hashtags around publications mentioned in Twitter and covered by Altmetric.com (2012–2016). *Nodes*: hashtags linked to more than 2000 publications in Altmetric.com. *Colors*: VOSviewer clustering result. *Edges*: publications in common between hashtags. Location of nodes in the map: VOSviewer clustering method

showed disciplinary differences in readership activity and topical interests [26.85].

26.6.2 Comparative Indicators

As presented in Table 26.1, comparative approaches use advanced indicators incorporating normalization features, such as field-normalized Mendeley indicators [26.90] or percentile-based indicators (e. g., Altmetric.com). The use of social media metrics as an evaluative device is the most problematic, since evaluative analysis requires higher levels of precision, validity and reliability. Moreover, the measurable concepts underlying most social media metrics are not clear [26.10]. Social media metrics for evaluative purposes can be distinguished in two groups: those that are conceptually similar to citations or peer review judgments (e. g., Mendeley or F1000Prime recommendations), and those that are not (e. g., Twitter or Facebook mentions).

Social Media Metrics Similar to Citations or Peer Review

Indicators such as readership in online reference managers (e. g., Mendeley or Zotero) and post-publication peer review platforms (e. g., F1000Prime, PeerJ or PubMed Commons) are conceptually close to citations and peer review judgments. Mendeley is used mainly by academic users [26.31, 32, 91], often in a pre-citation context [26.2]. Thus, both readership and citations may capture dimensions of *scientific influence*. Readership and citations are moderately correlated [26.55, 63, 92–94], more than other social media metrics [26.60, 95]. This suggests the potential relevance of Mendeley readership indicators as surrogates for citation-based indicators. This stronger correlation has encouraged field normalization of these indicators similar to citation indicators [26.90, 96], thereby opening the door to their use in more evaluative contexts. However, although close, citation and readership are still different. As argued by *Costas* et al. [26.74], the existence of two related but different metrics competing to capture the same concept may create conflicts (e. g., when one of the indicators points to high performance and the other to low performance). Given the higher engagement of an author citing a document in contrast to a Mendeley user saving a document [26.2], it is reasonable to argue that a citation is more valuable than a Mendeley readership. However, as argued by *Costas* et al. [26.58, 74], readership counts in Mendeley may be more meaningful than perfunctory citations [26.54]. This suggests that if the counts in Mendeley would include more qualitative aspects (e. g., indications of the time spent by the users in a given publication, or whether the users have made comments, notes, highlighted passages, appraised the text, etc.), the readership counts might be more informative in an evaluation context [26.97].

Other indicators for evaluative contexts include F1000Prime recommendations of publications provided by high-level appointed *experts*. This is a form of post-peer review evaluation, and these indicators are potentially interesting for quality judgment. However, they have two disadvantages. The first is the low number of publications reviewed and recommended in these services [26.65, 98], and the second is the weak correlation between these indicators and citation indicators [26.62, 65, 99], suggesting that they are related but not interchangeable indicators.

Social Media Metrics Dissimilar to Citations or Peer Review

Social media metrics, unlike citations or peer review, are not clearly related to scientific performance. Nevertheless, despite this limitation, some of these indicators have been proposed for evaluation. Indicators based on the h-index formula have been suggested (e. g., T-factor, see [26.100]; T-index [26.101]), as well as indicators inspired by the impact factor (Twimpact factor [26.102]), implicitly suggesting some straightforward comparability among them. Social media metrics do not relate directly to *scientific performance* (i. e., scientific impact or quality), but they may be related to *societal impact* [26.103]. However, even the concept of societal impact is quite nebulous and not easy to grasp. As a result, the jury is still out on the question of whether social media metrics are useful for research evaluation purposes.

To be useful for evaluation, most social media metrics must be conceptualized beyond the traditional research evaluation approaches. Thus, social media metrics may be relevant for evaluating the social media engagement of universities [26.44] or the public understanding of or engagement with science of different social media communities. From a policymaker perspective, social media metrics may also be used to evaluate *scientific literacy* among social media communities.

26.7 Prospects for Social Media Metrics in Research Evaluation

In the previous sections, we discussed the main characteristics, issues and practical possibilities related to social media metrics for research evaluation and management. Most social media metrics do not currently have practical application in the more traditional research evaluation approaches (i. e., those that would typically be based on peer review or citation analysis), perhaps with the exception of Mendeley and F1000Prime reviews. Therefore, the potential relevance of these indicators as scientific evaluative devices is still uncertain.

In this section, we take a more prospective (reflexive) perspective, in which we try to discuss and conceptualize potential (alternative) evaluative applications of social media metrics based on a fundamental understanding of their social media nature. We introduce more innovative perspectives on how different social media metrics could be used for new forms of evaluation. For example, a research organization that wishes to increase its visibility on Twitter as a means of expanding its social media visibility among broader communities of attention may use indicators such as PP(tw1) and communities-of-attention analysis to assess the realization of such an aim.

26.7.1 Understanding the Nature of Social Media Metrics for Research Evaluation

Current research evaluation methods do not focus on communication by social media, and instead are focused on the scholarly dimensions (although they are usually biased toward journal publications). Based on this dichotomy, we can introduce a novel approach for consideration of social media metrics. This perspective is related to the *foci* of the indicators. The *foci* of the different social media metrics can be determined based either on the aims of the platform (e. g., Twitter, Facebook have a purely social media focus) or on the nature of the indicator that is produced (e. g., the number followers in ResearchGate is a social media indicator, while the number of citations provided in the same platform could be seen as a scholarly indicator). Thus, we distinguish social media metrics with a stronger *social media focus* from social media metrics with a stronger *scholarly focus*. As *social media focus*, we understand the orientation of the tools, platforms, data and indicators that capture the interactions, sharing and exchange of information, ideas, messages, news, objects, etc. among diverse (online) users, and not necessarily restricted to scholarly users. As *scholarly focus*, we refer to those tools, platforms, data and indicators that are more oriented toward the management, analysis and evaluation of scholarly objects, entities and activities. Thus bibliometrics, citations and peer review can be considered as fundamentally having a scholarly focus.

Figure 26.8 illustrates the different foci of the most important bibliometric and social media metrics arranged in four quadrants based on their scholarly or social media focus. In the bottom-right part of the figure, we find the evaluative bibliometric and peer review indicators (represented by the databases Scopus and WoS and peers evaluating papers) with a strong scholarly focus (and low social media focus). In the top-left quadrant we find the platforms with the strongest social media focus (e. g., Twitter, Facebook, LinkedIn or Stack Exchange Q&A). These tools allow for the interaction and exchange of information among their users, but none of them have a genuinely scholarly focus (although the realm of social media metrics would confine itself to the interaction between these tools and scholarly objects). They have the largest distance from the scholarly-focused indicators. The main reason for this distance lies in the open, multipurpose and heterogeneous character of these platforms. Anyone can create a profile on Twitter, Facebook or LinkedIn and tweet or mention a scientific publication. Acts derived from these platforms, as argued in *Haustein* et al. [26.2], are driven by norms substantially different from those implicated in the act of citing (or peer review of) a publication.

In the bottom-right quadrant, in addition to the traditional bibliometrics (e. g., based on Scopus or Web of Science) and peer review, we also find F1000Prime recommendations and Mendeley readerships [26.31, 32, 91, 104–106], both with a reasonably strong scholarly focus (both are used mostly by scholars and are about scholarly outputs), although they also have some social media focus (e. g., both are user-generated, and interactions among users and outputs are possible). Wikipedia citations, while different from those found in scholarly publications (in theory, any person can write citations in a Wikipedia entry, although with some supervision), can still be considered similar enough to scholarly citations to be included in this quadrant.

In the top-right quadrant are platforms that combine both a strong social media and scholarly focus, such as ResearchGate and Academia.edu. These platforms are multipurpose, and their indicators are quite varied.

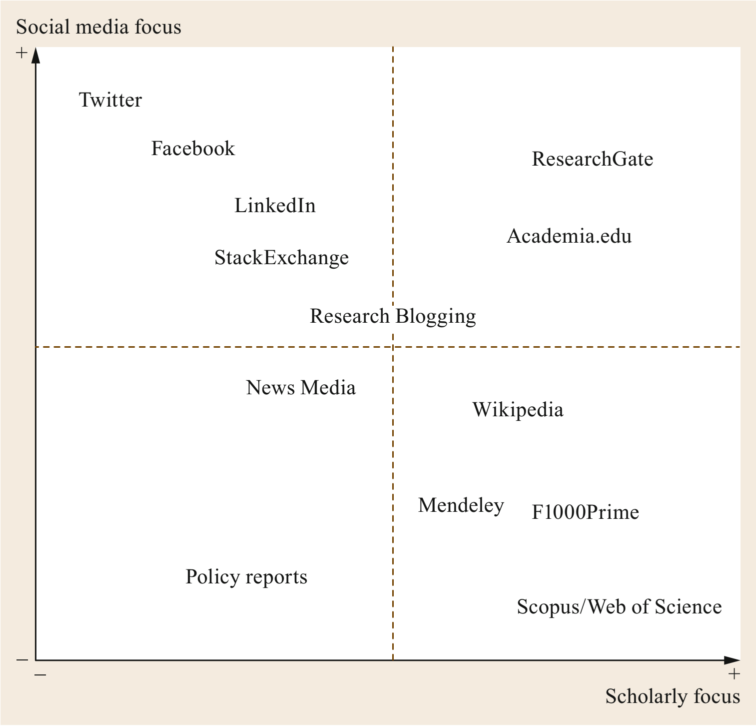


Fig. 26.8 Metrics characterized by their focus: social media or scholarly

These indicators can be grouped into those with a social media focus (e. g., the follower counts of scholars, number of endorsements, counts of Q&As on ResearchGate or the profile visits and mentions on Academia.edu) and those with a more scholarly focus (e. g., the counts of publications or citations, downloads and views on ResearchGate or Academia.edu). The RG Score combines elements from both these social media and scholarly foci into a single indicator, thus suggesting its potential unreliability.

In the bottom-left quadrant we find indicators that do not necessarily have either a social media or scholarly focus. An example is citations from policy documents (currently collected by Altmetric.com). Policy citations are of course relevant from several perspectives (e. g., policy impact, societal impact), but they are not created under the same norms as scholarly citations. Moreover, they do not have a social media focus (i. e., different types of users are not entitled to interact with the scholarly material discussed in the policy document). This calls into question whether policy document citations can be considered social media metrics at all.

In the center of the graph (Fig. 26.8) are mentions in blogs and news media. The central position of these indicators is explained by the fact that bloggers and science journalists could use scientific objects to support their arguments in their blog posts or news items and, as argued in *Haustein* et al. [26.2], could be driven by "similar *norms* as scholars", although not necessarily the same. Thus, these indicators would represent a bridge between the scholarly and social media foci.

26.7.2 Proposing Alternative Forms of Research Evaluation Based on Social Media Metrics

Based on the previous model, indicators with a stronger scholarly orientation would be more suitable for research evaluation (comparable to how citations and peer review are used). Thus, Mendeley readership and F1000Prime recommendations, and to some extent Wikipedia citations as well, could be seen as new tools for evaluating research [26.97]. As the social media focus of the indicators increases, one should consider how this would influence the evaluation (e. g., how nonacademic users in Mendeley could affect the indicators or how Wikipedia citations could be biased by nonacademic Wikipedia authors). Those social media metrics are more difficult to incorporate into the more traditional scholarly evaluations. However, social media metrics capture interactions between social media users and scientific objects. The relevance of social media activities is expanding in many walks of life, particularly in the dissemination of ideas, awareness and discussion of current issues, or sharing information, news and content. Many scholars, universities and scholarly organizations are mindful of their presence and image on these platforms. It is therefore not unreasonable to claim that the social media reception of scholarly objects can be seen as a nontrivial aspect of scientific communication. Monitoring the coverage, presence and reception of scientific objects on social media can then be seen as a novel element in research evaluation. The focus would not be on the scholarly impact or quality of the production of a research unit, but rather on the social media reception of its outputs.

New evaluations would include questions such as *How is the output of my university being discussed on Twitter? Are my publications visible among the relevant communities of attention? Do these communities engage with the publications? Is the social media reception and engagement of my output positive? Are the scholars of my unit active on social media? Do they contribute to disseminate their research and engage with broader communities to explain, expand or clarify their work? How are the social media communication strategies at the university working? etc.*

Table 26.3 Conceptualization of new social media metrics applications

Social media dimension	Example indicators (for a given research unit)
Coverage and presence of scholarly objects on social media	No. of publications mentioned on Twitter, Facebook etc. No. of scholars with a Twitter account Growth in % of publications mentioned on Twitter
Reception and attention on social media	No. of tweets to a given publication No. of tweets to a given publication with some degree of engagement No. of tweets to publications from highly followed tweeters
Engagement of social media users with scholarly objects	No. of tweets to a given publication containing comments, hashtags or remarks from the users
Communities of attention around scholarly objects	No. of tweeters tweeting the publications of the unit No. of highly followed tweeters tweeting the publications of the unit
Landscapes of social media attention around scholarly objects	No. of tweets to the outputs from the different fields of activity of the unit No. of tweets to outputs of the unit from social media users from different countries

Clearly, the questions above are new, and they may not be relevant for many research managers, but if social media matters, then social media metrics also matter. From this point of view, it is possible to conceptualize novel forms of research evaluation based on social media metrics. Table 26.3 summarizes (not exhaustively) some of the dimensions and indicators that can be considered in this social media evaluation of scientific objects of a given research unit.

26.8 Concluding Remarks

This chapter has brought together three different strands of literature:

1. The development of principles for good and responsible use of metrics in research assessments and post-publication evaluations
2. The technical literature on social media metrics and altmetrics
3. The literature about the conceptual meaning of social media metrics.

Thus, the chapter does not cover all forms of alternative research evaluations. For example, the increasing need for sustainable data infrastructure around data sets and the need to standardize the citation of data sets falls outside the scope of this chapter, although it is clearly of the utmost importance for the future of research evaluation. The need for data sharing and availability according to the findability, accessibility, interoperability and reusability (FAIR) principles requires a separate chapter. We have also not dealt with the interesting challenges that will be presented by the development of cloud computing in the context of research instruments and infrastructures for the conduct of research evaluations in the next decades. Nevertheless, by focusing on the novel measurement approaches that have developed as a result of the shift in research activities to the web, we hope the chapter has made clear how these data and indicators can be applied for practical purposes (and also how not to use them).

Our main proposal is to define the metrics formerly known as altmetrics primarily on the basis of their origin: as data and indicators of social media use, reception and impact in the context of academia. This both restricts and enables their use in research evaluations. Social media plays an important role in scientific and scholarly communication. It enables a faster distribution of data sets and preliminary results, and a greater level of access to formal research publications. It would therefore make sense to include this dimension of social media activity in research assessments whenever scientific communication is deemed relevant (of course this is not up to metrics experts to decide). We have sketched the outlines of such applications and have indicated the technical and conceptual challenges that need to be addressed.

Second, we propose to hold social media metrics accountable to the same principles of responsible metrics as are deemed valid for all performance metrics. As will be clear, although many social media indicators are easily available, they often fail with respect to transparency and openness. We find this ironic, given the original intent of social media metrics to open up the process of research evaluation.

A recent paper discussed the application of the 10 principles of the Leiden Manifesto for research metrics to social media metrics [26.107]. Like other metrics, social media metrics should only be used within the framework of *informed peer review*, and advanced normalized indicators are seen as preferred. The context of the research unit under evaluation should be taken into account. The use of altmetrics data should be transparent and freely accessible. As with traditional bibliometric indicators, false concreteness should be avoided. Systemic effects must be taken into consideration, and this may be more urgent for social media indicators since they are more easily gameable than citation indicators.

The currently developed principles for responsible metrics, therefore, do not need to be changed in order to be valid for social media metrics. But a large number of social media metrics seem to fail some of the principles, in particular, ironically, concerning the requirements of transparency, openness and manipulability. To address this, we may need a next-generation data infrastructure for social media metrics. Lastly, we propose discarding the term *altmetrics* and systematically starting to speak about specific *social media metrics* [26.34, 35], or even more generally, about *social media studies of science* [26.23, 81]. This then leaves sufficient space to develop new forms of indicators for scholarly objects (including publications, data sets and code, as well as scholars, scholarly organizations, etc.) and the use of research without conflating them with social media indicators.

Acknowledgments. The authors would like to thank Ludo Waltman for his critical comments on this chapter.

We also acknowledge partial funding from the South African DST-NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy (SciSTIP), the Centre for Research Quality and Policy Impact Studies (R-Quest; https://www.r-quest.no/) and the KNOWSCIENCE project (funded by the Riksbankens Jubileumsfond (RJ), https://www.fek.lu.se/en/research/research-groups/knowscience).

References

26.1 J. Priem, D. Taraborelli, P. Groth, C. Neylon: Altmetrics: A manifesto, http://altmetrics.org/manifesto/ (2010)

26.2 S. Haustein, T.D. Bowman, R. Costas: Interpreting altmetrics: Viewing acts on social media through the lens of citation and social theories. In: *Theories of Informetrics and Scholarly Communication*, ed. by C.R. Sugimoto (De Gruyter, Berlin 2016) pp. 372–406

26.3 T. Almind, P. Ingwersen: Informetric analyses on the world wide web: Methodological approaches to "webometrics", J. Doc. **53**, 404–426 (1997)

26.4 B. Cronin, H.W. Snyder, H. Rosenbaum, A. Martinson, E. Callahan: Invoked on the web, J. Am. Soc. Inf. Sci. **49**(14), 1319–1328 (1998)

26.5 R. Rousseau: Sitations: An exploratory study, Cybermetrics **1**(1), 1–7 (1997), http://hdl.handle.net/10261/170110

26.6 J. Priem, H. Piwowar, B. Hemminger: Altmetrics in the wild: Using social media to explore scholarly impact, arXiv:1203.4745v1 (2012) [cs.DL]

26.7 J. Priem, B.H. Hemminger: Scientometrics 2.0: New metrics of scholarly impact on the social web, First Monday (2010), https://doi.org/10.5210/fm.v15i7.2874

26.8 J. Priem: Scholarship: Beyond the paper, Nature **495**(7442), 437–440 (2013)

26.9 P. Wouters, M. Thelwall, K. Kousha, L. Waltman, S. De Rijcke, A. Rushforth, T. Franssen: The metric tide: Literature review (supplementary report I to the independent review of the role of metrics in research assessment and management), Hefce (2015), https://doi.org/10.13140/RG.2.1.5066.3520

26.10 P. Wouters, R. Costas: Users, narcissism and control—Tracking the impact of scholarly publications in the 21st century, Utrecht: SURFfoundation, http://research-acumen.eu/wp-content/uploads/Users-narcissism-and-control.pdf (2012)

26.11 National Information Standards Organization: Outputs of the NISO Alternative Assessment Metrics Project (2016), https://groups.niso.org/apps/group_public/download.php/17091/NISO+RP-25-2016+Outputs+of+the+NISO+Alternative+Assessment+Project.pdf

26.12 J. Wilsdon, L. Allen, E. Belfiore, P. Campbell, S. Curry, S. Hill, R. Jones, R. Kain, S. Kerridge, M. Thelwall, J. Tinkler, I. Viney, P. Wouters, J. Hill, B. Johnson: *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management* (HEFCE, Stoke Gifford 2015), https://doi.org/10.13140/RG.2.1.4929.1363

26.13 D. Hicks, P. Wouters, L. Waltman, S. de Rijcke, I. Rafols: The Leiden manifesto for research metrics, Nature **520**, 429–431 (2015)

26.14 San Francisco Declaration on Research Assessment (DORA): http://www.ascb.org/dora/ (2013)

26.15 W. Glänzel: The need for standards in bibliometric research and technology, Scientometrics **35**(2), 167–176 (1996)

26.16 H.F. Moed: *Citation Analysis in Research Evaluation*, Vol. 9 (Springer, Dordrecht 2005)

26.17 L. Leydesdorff, P. Wouters, L. Bornmann: Professional and citizen bibliometrics: Complementarities and ambivalences in the development and use of indicators—A state-of-the-art report, Scientometrics **109**(3), 2129–2150 (2016)

26.18 M. Thelwall: Introduction to webometrics: Quantitative web research for the social sciences, Synth. Lect. Inf. Concepts Retr. Serv. **1**(1), 1–116 (2009)

26.19 P. Wouters, W. Glänzel, J. Gläser, I. Rafols: The dilemmas of performance indicators of individual researchers: An urgent debate in bibliometrics, ISSI Newsletter **9**(3), 48–53 (2013)

26.20 P. Wouters, J. Bar-Ilan, M. Thelwall, I.F. Aguillo, Ü. Must, F. Havemann, J. Schneider: *ACUMEN Final Report* (ACUMEN Consortium, Brussels 2014), http://research-acumen.eu/wp-content/uploads/ACUMEN-Guidelines-Portfolio-TOC-document-v13.pdf

26.21 H.F. Moed: Altmetrics as traces of the computerization of the research process. In: *Theories of Informetrics and Scholarly Communication*, ed. by C.R. Sugimoto (De Gruyter, Berlin, Boston 2016)

26.22 S. Haustein: Grand challenges in altmetrics: Heterogeneity, data quality and dependencies, Scientometrics **108**(1), 413–423 (2016)

26.23 R. Costas: Towards the social media studies of science: Social media metrics, present and future, Bibl. An. Investig. **13**(1), 1–5 (2017)

26.24 H. Shema, J. Bar-Ilan, M. Thelwall: Do blog citations correlate with a higher number of future citations? Research blogs as a potential source for alternative metrics, J. Assoc. Inf. Sci. Technol. **65**(5), 1018–1027 (2014)

26.25 I. Rowlands, D. Nicholas, B. Russell, N. Canty, A. Watkinson: Social media use in the research workflow, Learn. Publ. **24**(3), 183–195 (2011)

26.26 E. Orduña-Malea, A. Martín-Martín, E.D. López-Cózar: ResearchGate como fuente de evaluación científica: Desvelando sus aplicaciones bibliométricas, Prof. Inf. **25**(2), 303–310 (2016)

26.27 M. Thelwall, K. Kousha: ResearchGate versus Google Scholar: Which finds more early citations?, Scientometrics (2017), https://doi.org/10.1007/s11192-017-2400-4

26.28 Z. Zahedi, R. Costas: General discussion of data quality challenges in social media metrics: Extensive comparison of four major altmetric data aggregators, PloS ONE **13**(5), e0197326 (2018)

26.29 P.B. Brandtzæg: Towards a unified media-user typology (MUT): A meta-analysis and review of the research literature on media-user typologies, Comput. Hum. Behav. **26**(5), 940–956 (2010)

26.30 S.J. Barnes, H. Bauer, M. Neumann, F. Huber: Segmenting cyberspace: A customer typology for the internet, Eur. J. Mark. **41**(1), 71–93 (2007)

26.31 S. Haustein, V. Larivière: Mendeley as a source of global readership by students and postdocs? Evaluating article usage by academic status. In: *35th Int. Assoc. Technol. Univ. Librar. (IATUL), 2–5 June 2014, Aalto Univ., Helsinki* (2014) pp. 1–10, http://docs.lib.purdue.edu/iatul/2014/altmetrics/2

26.32 Z. Zahedi, R. Costas, P. Wouters: Assessing the impact of publications saved by Mendeley users: Is there any different pattern among users? In: *35th Int. Assoc. Technol. Univ. Librar. (IATUL), 2–5 June 2014, Aalto Univ., Helsinki* (2014) pp. 1–12, http://docs.lib.purdue.edu/iatul/2014/altmetrics/4

26.33 E. Mohammadi, M. Thelwall: Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows, J. Assoc. Inf. Sci. Technol. **65**(8), 1627–1638 (2014)

26.34 S. Haustein, T.D. Bowman, R. Costas: Communities of attention around journal papers: Who is tweeting about scientific publications? In: *Soc. Media Soc. 2015 Int. Conf., Toronto* (2015) pp. 4–5

26.35 S. Haustein, T.D. Bowman, R. Costas: When is an article actually published? An analysis of online availability, publication, and indexation dates. In: *15th Int. Conf. Scientom. Informetr. (ISSI)* (2015) pp. 1170–1179, arXiv:1505.00796v1

26.36 S. Haustein, R. Costas: Identifying Twitter audiences: Who is tweeting about scientific papers? In: *ASISandT SIG/MET Metrics 2015 Workshop Identifying, St. Louis* (2015) pp. 9–11

26.37 R. Costas, J. van Honk, T. Franssen: Identifying scholars on Twitter: Opening the path to the social media studies of science. In: *4 AM Conf.* (2017), http://altmetricsconference.com/register-for-4am-in-toronto/

26.38 R. Costas, J. van Honk, T. Franssen: Scholars on Twitter: Who and how many are they? In: *16th Int. Conf. Scientom. Informetr. (ISSI), Wuhan, Wuhan Univ.* (2017) pp. 224–235, https://arxiv.org/abs/1712.05667

26.39 Q. Ke, Y. Ahn, C.R. Sugimoto: A systematic identification and analysis of scientists on Twitter, PLOS ONE **12**(4), 1–28 (2016)

26.40 D.M. Boyd, N.B. Ellison: Social network sites: Definition, history, and scholarship, J. Comput. Mediat. Commun. **13**(1), 210–230 (2007)

26.41 R.D. Putnam: *Bowling Alone: The Collapse and Revival of American Community* (Simon Schuster, New York 2000)

26.42 M. Hofer, V. Aubert: Perceived bridging and bonding social capital on Twitter: Differentiating between followers and followees, Comput. Hum. Behav. **29**(6), 2134–2142 (2013)

26.43 M. Kaigo: Social media usage during disasters and social capital: Twitter and the great east Japan earthquake, Keio Commun. Rev. **34**, 19–35 (2012)

26.44 N. Robinson-Garcia, T.N. van Leeuwen, I. Rafols: Using altmetrics for contextualised mapping of societal impact: From hits to networks, Sci. Publ. Policy **45**(6), 815–862 (2017)

26.45 C. Hank, C.R. Sugimoto, A. Tsou, J. Pomerantz: Faculty and student interactions via Facebook: Policies, preferences, and practices, Inf. Technol. **56**(5), 216–223 (2014)

26.46 S. Haustein, R. Costas, V. Larivière: Characterizing social media metrics of scholarly papers: The effect of document properties and collaboration patterns, PLoS ONE **10**(3), e0120495 (2015)

26.47 H. Shema, J. Bar-Ilan, M. Thelwall: How is research blogged? A content analysis approach, J. Assoc. Inf. Sci. Technol. **66**(6), 1136–1149 (2015)

26.48 Y. Gingras: Criteria for evaluating indicators. In: *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*, ed. by B. Cronin, C.R. Sugimoto (MIT Press, Cambridge 2014) pp. 108–125

26.49 R.E. Rice, C.L. Borgman, D. Bednarski, P.J. Hart: Journal-to-journal citation data: Issues of validity and reliability, Scientometrics **15**(3/4), 257–282 (1989)

26.50 A.J. Nederhof: The validity and reliability of evaluation scholarly performance. In: *Handbook of Quantitative Studies of Science and Technology*, ed. by A.F.J. van Raan (Elsevier, Amsterdam 1988) pp. 193–228

26.51 L. Bornmann, H.-D. Daniel: What do citation counts measure? A review of studies on citing behavior, J. Doc. (2008), https://doi.org/10.1108/00220410810844150

26.52 M.H. MacRoberts, B.R. MacRoberts: Citation analysis and the science policy arena, Trends Biochem. Sci. **14**(1), 8–12 (1989)

26.53 M.H. MacRoberts, B.R. MacRoberts: The mismeasure of science: Citation analysis, J. Assoc. Inf. Sci. Technol. (2017), https://doi.org/10.1002/asi.23970

26.54 J. Nicolaisen: Citation analysis, Annu. Rev. Inf. Sci. Technol. **41**, 609–641 (2007)

26.55 N. Maflahi, M. Thelwall: When are readership counts as useful as citation counts? Scopus versus Mendeley for LIS journals, J. Assoc. Inf. Sci. Technol. **67**(1), 191–199 (2016)

26.56 M. Thelwall: Are Mendeley reader counts high enough for research evaluations when articles are published?, Aslib J. Inf. Manag. **69**(2), 174–183 (2017)

26.57 Z. Zahedi, R. Costas, P. Wouters: Mendeley readership as a filtering tool to identify highly cited

publications, J. Assoc. Inf. Sci. Technol. **68**(10), 2511–2521 (2017)
26.58 R. Costas, S. Haustein, Z. Zahedi: Larivière: Exploring paths for the normalization of altmetrics: Applying the characteristic scores and scales. In: *Altmetrics16. Moving Beyond Counts: Integrating Context, Bucharest* (2016) pp. 1–8
26.59 M. Thelwall, K. Kousha: Web indicators for research evaluation. Part 2: Social media metrics, Prof. Inf. **24**(5), 607 (2015)
26.60 R. Costas, Z. Zahedi, P. Wouters: Do altmetrics correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective, J. Assoc. Inf. Sci. Technol. **66**(10), 2003–2019 (2015)
26.61 M. Nardo, M. Saisana, A. Saltelli, S. Tarantola, A. Hoffman, E. Giovannini: *Handbook on Constructing Composite Indicators Methodology and User Guide*, OECD Statistics Working Papers (OECD, Paris 2005)
26.62 X. Li, M. Thelwall: F1000, Mendeley and traditional bibliometric indicators. In: *Proc. 17th Int. Conf. Sci. Technol. Indicators*, Vol. 3 (2012) pp. 451–551
26.63 Z. Zahedi, R. Costas, P. Wouters: How well developed are altmetrics? A cross-disciplinary analysis of the presence of "alternative metrics" in scientific publications, Scientometrics **101**(2), 1491–1513 (2014)
26.64 R. Haunschild, L. Bornmann: F1000Prime: An analysis of discipline-specific reader data from Mendeley, F1000Research **41**, 1–5 (2015)
26.65 L. Waltman, R. Costas: F1000 recommendations as a potential new data source for research evaluation: A comparison with citations, J. Assoc. Inf. Sci. Technol. **65**(3), 433–445 (2014)
26.66 M. Thelwall, S. Haustein, V. Larivière, C.R. Sugimoto: Do altmetrics work? Twitter and ten other social web services, PLoS ONE **8**(5), e64841 (2013)
26.67 K. Holmberg: The impact of retweeting on altmetrics. In: *Altmetrics14: Expand. Impacts Metrics ACM Web Sci. Conf. Workshop, Bloomintong* (2014) pp. 2–4, https://doi.org/10.6084/m9.figshare.1037555
26.68 C.R. Sugimoto, S. Work, V. Larivière, S. Haustein: Scholarly use of social media and altmetrics: A review of the literature, J. Assoc. Inf. Sci. Technol. **68**(9), 2037–2062 (2017)
26.69 N. Mallig: A relational database for bibliometric analysis, J. Informetr. **4**(4), 564–580 (2010)
26.70 F. Narin: *Evaluative Bibliometrics: The Use of Publication and Citation Anlaysis in the Evaluation of Scientific Activity* (Computer Horizons, Washington 1976)
26.71 T. van Leeuwen: Descriptive versus evaluative bibliometrics. In: *Handbook of Quantitative Science and Technology Research*, ed. by H.F. Moed, W. Glänzel, U. Schmoch (Springer, Dordrecht 2004) pp. 373–388
26.72 L. Waltman, N.J. Van Eck: The need for contextualized scientometric analysis: An opinion paper. In: *STI Conference 2016. Peripheries, Frontiers and Beyond*, ed. by I. Ràfols, J. Molas-Gallart, E. Castro-Martínez, R. Woolley (Universitat Politècnica de València, Valencia 2016) pp. 1–9
26.73 R. Costas, Z. Zahedi, P. Wouters: The thematic orientation of publications mentioned on social media large-scale disciplinary comparison of social media metrics with citations, Aslib J. Inf. Manag. **67**(3), 260–288 (2015)
26.74 R. Costas, A. Perianes-Rodríguez, J. Ruiz-Castillo: On the quest for currencies of science: Field "exchange rates" for citations and Mendeley readership, Aslib J. Inf. Manag. **69**(5), 557–575 (2017)
26.75 M. Thelwall: Three practical field normalised alternative indicator formulae for research evaluation, J. Inf. **11**(1), 128–151 (2016)
26.76 G. Colavizza, M. Franceschet: Clustering citation histories in the Physical Review, J. Informetr. **10**(4), 1037–1051 (2016)
26.77 L. Waltman, N.J. van Eck: Field-normalized citation impact indicators and the choice of an appropriate counting method, J. Informetr. **9**(4), 872–894 (2015)
26.78 N. Maflahi, M. Thelwall: How quickly do publications get read? The evolution of mendeley reader counts for new articles, J. Assoc. Inf. Sci. Technol. (2017), https://doi.org/10.1002/asi.23909
26.79 L. Waltman, N.J. Van Eck: A new methodology for constructing a publication-level classification system of science, J. Am. Soc. Inf. Sci. Technol. **63**(12), 2378–2392 (2012)
26.80 N.J. van Eck, L. Waltman: Software survey: VOSviewer, a computer program for bibliometric mapping, Scientometrics **84**(2), 523–538 (2010)
26.81 R. Costas, J. Van Honk, C. Calero-Medina, Z. Zahedi: Discussing practical applications for altmetrics: Development of "social media profiles" for African, European and North American publications. In: *22nd Int. Conf. Sci. Technol. Indicators (STI), Paris* (2017), https://sti2017.paris/full-programme/
26.82 J. van Honk, R. Costas: Integrating context in Twitter metrics: Preliminary investigation on the possibilities of hashtags as an altmetric resource. In: *Altmetrics16. Moving Beyond Counts: Integrating Context, Bucharest* (2016) pp. 1–7, http://altmetrics.org/wp-content/uploads/2016/09/altmetrics16_paper_5.pdf
26.83 R. Haunschild, L. Bornmann, L. Leydesdorff: Networks of reader and country status: An analysis of Mendeley reader statistics, PeerJ Comput. Sci. **1**, e32 (2015)
26.84 P. Kraker, C. Schlögl, K. Jack, S. Lindstaedt: Visualization of co-readership patterns from an online reference management system, J. Informetr. **9**(1), 169–182 (2015)
26.85 Z. Zahedi, N. Van Eck: Exploring topics of interest of Mendeley users, J. Altmetrics **1**(1), 5 (2018)
26.86 Z. Zahedi, N.J. Van Eck: Visualizing readership activity of Mendeley users using VOSviewer. In: *Altmetrics14: Expand. Impacts Metrics ACM Web Sci. Conf. 2014 Altmetrics Workshop, Indiana Univ.*

(Figshare, Indiana 2014), https://doi.org/10.6084/m9.figshare.1041770
26.87 J. Jiang, D. He, C. Ni: Social reference. In: *Proc. 11th Annu. Int. ACM/IEEE Joint Conf. Digital Librar. – JCDL'11*, Vol. 11 (ACM, New York 2011) p. 401
26.88 K.W. Boyack, R. Klavans: Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately?, J. Am. Soc. Inf. Sci. Technol. **61**(12), 2389–2404 (2010)
26.89 H. Small: Co-citation in the scientific literature: A new measure of the relationship between two documents, J. Am. Soc. Inf. Sci. **24**(4), 265–269 (1973)
26.90 R. Haunschild, L. Bornmann: Normalization of Mendeley reader counts for impact assessment, J. Informetr. **10**(1), 62–73 (2016)
26.91 E. Mohammadi, M. Thelwall, S. Haustein, V. Larivière: Who reads research articles? An altmetrics analysis of Mendeley user categories, J. Assoc. Inf. Sci. Technol. **66**(9), 1832–1846 (2015)
26.92 J. Bar-Ilan: Astrophysics publications on arXiv, Scopus and Mendeley: A case study, Scientometrics **100**(1), 217–225 (2014)
26.93 M. Thelwall, P. Sud: Mendeley readership counts: An investigation of temporal and disciplinary differences, J. Assoc. Inf. Sci. Technol. **67**(12), 3036–3050 (2016)
26.94 D. Torres-Salinas, Á. Cabezas-Clavijo, E. Jiménez-Contreras: Altmetrics: New indicators for scientific communication in Web 2.0, Comunicar **21**(41), 53–60 (2013)
26.95 S. Haustein, I. Peters, C.R. Sugimoto, M. Thelwall, V. Larivière: Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature, J. Assoc. Inf. Sci. Technol. **65**(4), 656–669 (2014)
26.96 L. Bornmann, R. Haunschild: Normalization of Mendeley reader impact on the reader- and paper-side: A comparison of the mean discipline normalized reader score (MDNRS) with the mean normalized reader score (MNRS) and bare reader counts, J. Informetr. **10**(3), 776–788 (2016)
26.97 Z. Zahedi: *Understanding the value of social media metrics for research evaluation*, Doctoral dissertation (Centre for Science & Technology Studies (CWTS), Leiden Univ., Leiden 2018)
26.98 L. Bornmann: Validity of altmetrics data for measuring societal impact: A study using data from Altmetric and F1000Prime, J. Informetr. **8**(4), 935–950 (2014)
26.99 L. Bornmann, L. Leydesdorff: The validation of (advanced) bibliometric indicators through peer assessments: A comparative study using data from InCites and F1000, J. Informetr. **7**(2), 286–291 (2013)
26.100 L. Bornmann, R. Haunschild: t factor: A metric for measuring impact on Twitter, Malays. J. Lib. Inf. Sci. **21**(2), 13–20 (2016)
26.101 H. Piwowar, J. Priem: Researcher-level altmetrics. In: *3:Am Conf., Bucharest* (2016), https://doi.org/10.6084/m9.figshare.3980091.v1
26.102 G. Eysenbach: Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact, J. Med. Internet Res. **13**(4), e123 (2011)
26.103 L. Bornmann: What is societal impact of research and how can it be assessed? A literature survey, J. Am. Soc. Inf. Sci. Technol. **64**(2), 217–233 (2013)
26.104 L. Bornmann, R. Haunschild: Which people use which scientific papers? An evaluation of data from F1000: Discovery Service para Universidad de La Rioja, J. Informetr. **9**(3), 477–487 (2015)
26.105 Z. Zahedi, R. Costas, V. Larivière, S. Haustein: On the relationships between bibliometric and altmetric indicators: The effect of discipline and density level. In: *The 2016 Altmetrics Workshop, Bucharest* (2016)
26.106 Z. Zahedi, S. Haustein: On the relationships between bibliographic characteristics of scientific documents and citation and Mendeley readership counts: A large-scale analysis of web of science publications, J. Informetr. **12**(1), 191–202 (2018)
26.107 L. Bornmann, R. Haunschild: To what extent does the Leiden manifesto also apply to altmetrics? A discussion of the manifesto against the background of research into altmetrics, Online Inf. Rev. **40**(4), 529–543 (2016)

Paul Wouters

Centre for Science and Technology Studies (CWTS)
Leiden University
Leiden, The Netherlands
p.f.wouters@cwts.leidenuniv.nl

Paul Wouters is Professor of Scientometrics and Director of the Centre for Science and Technology Studies (CWTS) at Leiden University (the Netherlands). He has published on the history of the Science Citation Index, on and in scientometrics, and on the way the criteria of scientific quality and relevance have been changed by the use of performance indicators. He was Principal Investigator for several European research consortia.

Zohreh Zahedi

Centre for Science and Technology Studies (CWTS)
Leiden University
Leiden, The Netherlands
z.zahedi.2@cwts.leidenuniv.nl

Zohreh Zahedi obtained her PhD degree at the Centre for Science and Technology Studies (CWTS) of Leiden University. Her PhD thesis centers around the value of using social media metrics for research evaluation. Her areas of expertise include social media metrics and scientometrics.

Rodrigo Costas

Centre for Science and Technology Studies (CWTS)
Leiden University
Leiden, The Netherlands
rcostas@cwts.leidenuniv.nl

Rodrigo Costas is Senior Researcher at the Centre for Science and Technology Studies (CWTS) at Leiden University (the Netherlands). Rodrigo is also an Extraordinary Associate Professor at the Centre for Research on Evaluation, Science and Technology of Stellenbosch University (South Africa). He holds a PhD in Library and Information Science from the CSIC in Spain. His areas of expertise include the fields of information science, scientometrics, and social media metrics.

27. Reviewing, Indicating, and Counting Books for Modern Research Evaluation Systems

Alesia Zuccala, Nicolas Robinson-García

In this chapter, we focus on the specialists who have helped to improve the conditions for book assessments in research evaluation exercises, with empirically based data and insights supporting their greater integration. Our review highlights the research carried out by four types of expert communities—the *monitors*, the *subject classifiers*, the *indexers*, and the *indicator constructionists*. Many challenges lie ahead for scholars affiliated with these communities, particularly the latter three. By acknowledging their unique yet interrelated roles, we show where the greatest potential is for both quantitative and qualitative indicator advancements in book-inclusive evaluation systems.

27.1 Evaluating Scholarly Books

Since antiquity, books have evolved remarkably. The earliest *books* were first carved onto clay tablets, and then painted onto papyrus scrolls. In China they were cut into a woodblock, and later, the Europeans printed full manuscripts with ink on paper. Now we have electronic books, or e-books, available online for download. In light of these transformations, seminal volumes have also been written about the history of the printing press [27.1], book publishing [27.2], and book classification systems [27.3], including new perspectives on the book in the digital era [27.4]. There is much to learn from this history in order to review, indicate, and count books for modern research evaluation systems.

A valuable starting point is recognizing that books, tightly bound for centuries, are somewhat paradoxical: the information they contain can have the power to liberate. Books have been and continue to be change agents in society [27.1]. They change the way that humans think and feel, remind us of our triumphs and follies, and can start a debate or incite a revolution. Some books are lauded; others are not. Some have even been banned from public consumption. Yet, all because of Gutenberg's printing press [27.1, p. 520]:

> A new kind of collaborative venture in data collection [was] set in motion even before laboratory facilities were built, or new observational instruments had been invented. The shift from script to print helps to explain why old theories were found wanting and new ones devised even before telescopes, microscopes, and scientific societies appeared. Gutenberg's invention not only preceded Galileo's tube; it was a more versatile data aid and affected a wider range of data. Some professors shunned controversy and withheld treatises from the press just as some refused at first to look through the telescope. But none failed to consult printed reference guides or preferred to have to copy out tables by hand. Whatever views were held concerning Aristotle, Ptolemy or Galen, whatever objections were posed against using vernaculars or courting publicity; printed maps, charts, and diagrams found rapid acceptance from all.

With this level of acceptance comes great responsibility on the part of evaluators. Books, in essence, capture the efforts of scholars concerned with various types of human endeavors [27.5]. Yet, for many years, the

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_27

evaluation community has focused on journal articles rather than books. Since the 1960s, the journal article has taken precedence as "a written and published report describing original research results" [27.6, p. 8]. In this regard, books and book publishers lag behind, even though they

> stand at a crucial crossroads in the production and distribution of knowledge in any society. They are in a position to decide what is *in* and what is *out* of the marketplace of ideas. [27.7, p. 14]

In addition to the journal article, this means that the book needs to be delineated or more clearly defined. In the simplest of terms, *Williams* et al. [27.8] note that "what differentiates a book from a periodical or long report" is that it qualifies for and has an ISBN. *Basili* and *Lanzillo* suggest that an authored book, or monograph, may be defined as [27.9, p. 162]:

> The product of an intense but wide-ranging, systematic and unified research examination of an area of study. Each element contributes to forming the complex of the work, which could not be successfully communicated through the publication of separate parts.

The monograph's purpose is to present

> what the scholar concludes is the truth about some set of historical events, the characteristics of some work of art or literature, or the biography of a historical figure, an artist, or a writer. [27.10]

Hence, with a series of scholarly monographs we can piece together the story of a research discipline—i. e., how it has evolved in different regions, over a specific time period, and within a particular interpretive community [27.11].

In recent decades, the research evaluation community has questioned the value of the book. Adding to this problem has been the decline in sales of scholarly monographs since the 1980s [27.2], including a shift on the part of some researchers toward publishing more journal articles [27.12]. Surrounding such publication practices there has also been a lack of stable methods and indicators available to properly assess the monograph's value, impact, or influence. Still, research evidence indicates that authored books prevail, and will continue to prevail, because they hold meaning for certain research communities, distinct from those observed in journal literature [27.13–15].

To evaluate scholarly books and account for the influence they have had on their readership, a balancing act is required. On one hand, the evaluative process should respect all that an authored book represents in *qualitative* terms, both to the writer and to his/her audience. Book reviews help maintain this respect for quality, since the process of reviewing can be at the same time descriptive, appreciative, and critical. On the other hand, emergent digital tools are now inspiring researchers to devise new ways of assigning symbolic forms of credit to them en masse. Google, for example, wants all books around the world to "stand up and be counted" [27.16]. Clarivate Analytics' decree has been less direct, though critics of the expanding commercial Book Citation Index℠ have much to say about the opportunities and limitations associated with "putting books back into the library" [27.17, 18].

In this chapter, we review some of the approaches taken thus far to evaluate books both qualitatively and quantitatively. Our focus is on the cluster of information specialists without whom the practice of research evaluation would not prosper:

1. *The monitors*
2. *The subject classifiers*
3. *The indexers*
4. *The indicator constructionists*.

This work also provides suggestions for future evaluation systems dedicated to safeguarding book-oriented research fields, so that they can continue to develop progressively. As with any guide for evaluation, the "crucial issue at stake is not whether scholars' practices change", but that the application of any specific tool of measurement "enhances research performance and scholarly progress in general" [27.19, p. 578].

27.2 The Monitors

A monitor may be described as someone who observes, keeps track of, or surveys the progress or quality of something over time. In this sense, many researchers have played monitoring roles for wider aspects of the research evaluation systems and for books as well. There is a need for monitors because they show us what is possible to evaluate, where data/information is lacking, and what could be improved upon in the future. While monitors often detect the potential for qualitative or quantitative indicators, they usually do not focus on developing them fully for formal use. They are the historical benefactors of our current system, having brought us to where we are today with the evaluation of books and book-oriented research fields.

Long before the creation of commercial book indexes—i. e., Clarivate Analytics' Book Citation Index℠ and Elsevier's Scopus—researchers were less interested in metrics for books, and focused more on the uses (or misuses) of published book reviews [27.20]. For librarians in particular, the book review was and still is considered a valuable aid for building book collections [27.21–23]. Within many library communities it has therefore become essential to study the review culture as a unique form of discourse, and to consider the merits of applying standards for reviewing [27.24–27]. However, with scholars also reading and making use of book reviews [27.28, 29], researchers further recognize that even though a review is not an original work, it can still transfer useful information and ideas. For instance, there is an expectation that a review based on a book published in history will appear in a history journal, but a review could be written about the same book and be published in a political science journal as well. *Lindholm-Romatschuk* [27.30] explains this transfer in terms of "intradisciplinary and interdisciplinary information flows" [27.30, p. 86].

Another critical stage in book-based evaluations took place in the 1970s, when researchers began to dissect book reviews using different methods of content analysis [27.31–34]. Most of this early work had to be done manually, using data sets of approximately 1000–2000 reviews. Although researchers today have better technologies for working with data, the first content-based studies marked the beginning of a positive trend toward an "informed sociology of the review process" [27.35, p. 114].

The research of *Champion* and *Morris* [27.32] and of *Bilhartz* [27.31] highlights the degree to which specific time periods have had an effect on review discourse. While book reviews of the 1960s tended to be *gentlemanly* and mostly favorable, those published in the late 1970s and 1980s, specifically for the field of history, increasingly devoted more space to "critiquing rather than simply summarizing a book's content" [27.31, p. 527]. *Bilhartz* [27.31] found specifically that reviewers of the 1970s "took a strong interest in originality of method". However, "more than in any previous decade", reviewers of the 1980s "expect[ed] histories to have a sharply focused and well-analyzed thesis" [27.31, pp. 527–528]. *Snizek* and *Fuhrman* [27.35] consequently hypothesized and later found that favorability in a published book review was significantly and positively correlated with both the age and experience of the reviewer.

Gradually, the monitoring phase shifted when information scientists decided to test quantitative techniques for assessing books. *Eugene Garfield* [27.36], creator of the first Science Citation Index, suggested that the creation of a book index would support the *biblio* or book-oriented side of *biblio*-metrics, but, in the absence of this tool, researchers turned to journal indexes in order to analyze book publication, citation, and book review counts [27.37–41]. Some scholars were working with books as distinct study objects [27.42, 43], while others wanted to give more credit to the book as the principal form of publication across the humanities and/or social sciences [27.44–46].

With this preliminary stage of book-oriented metrics came the notion that both the humanities and social sciences were at risk of being poorly represented and unfairly assessed [27.47]. What some of the first *biblio*-metricians did, essentially, was to bring to light critical questions about how to assess the social sciences and humanities, primarily because they can be more theory-oriented, and progress more slowly than the sciences [27.48]. Emphasis was placed on scholars from certain disciplines who might be sharing information using media other than journals (i. e., books!), or contributing to local outlets, including those directed to a non-scholar public [27.38, 49]. This led to a significant debate concerning the development and use of alternative databases, like Google Books [27.50], or relying more seriously on the open access movement and institutional repositories [27.51–54].

Today, the evaluation community can turn to the Book Citation Index℠, and researchers also have the possibility of assessing publication and citation counts for books using the Elsevier Scopus database. But commercial databases of this nature are a type of *library*, and as researchers subscribe to or become patrons of these unique digital *libraries*, it will be increasingly necessary for them to understand how books are categorized and indexed. This cannot be taken for granted, and in fact with the first Science Citation Index, it was also a primary issue.

27.3 The Subject Classifiers

When *Eugene Garfield* [27.55] first conceived of *a new dimension in documentation through the association of ideas* (i. e., the Science Citation Index), he reflected on the following [27.55, p. 108]:

> If one considers the book as the macro unit of thought and the periodical article the micro unit of thought, then the citation index in some respects deals in the submicro or molecular unit of thought. It is here that most indexes are inadequate, because the scientist is quite often concerned with a particular idea rather than with a complete concept. *Thought* indexes can be extremely useful if they are properly conceived and developed.... One of the basic difficulties is to build subject indexes that can anticipate the infinite number of possible approaches the scientist may require.

Clearly, subject areas of *thought* were foremost in Garfield's mind, and like the indexes created for journals, subject-based catalogs for books were developed primarily for retrieval purposes. Unlike journals, subject classifications have not yet been used in the development of metric evaluations. However, early monitoring pertaining to book publication and citation counts suggests that a book citation index might indeed be used for this purpose. It is therefore useful to compare subject classes/categories designed for journals with those conceived for books, although a history of the latter is older.

Throughout the 16th and 17th centuries, books held in traditional library stacks were not open to the general public, and were available only to special users. Once opened, it became important to position books on the shelves so that patrons could locate them in relative terms. Melville Dewey, inventor of the *Dewey Decimal Classification System*, recognized that books could be grouped together on the basis of similar topics. In 1876, he published the first classification and subject index for books and pamphlets. Several editions of his classification system were published in both English and French (i. e., the French *Classification Decimal*), including an abridged edition, a library edition, and a bibliographic edition, which later became known as the *Universal Decimal Classification*.

After Dewey's death in 1931, the editorship of his classification volumes fell to the Library of Congress. By the time the 16th and 17th editions were published, Dewey's system had been widely adopted by general libraries, but a new Library of Congress Classification (LCC) had also been devised for larger, research-oriented libraries. Both the Dewey Decimal Classification System (DDC) and the Library of Congress Classification (LCC) system possess comparable subject codes and descriptors [27.56]. The LCC, however, adds an extra Cutter number (a Cutter number refers to the system developed by Charles Ammi Cutter, who invented the *Cutter Expansive Classification System*), which is used to represent a book's author, title, or organization name. While the DDC and LCC are the predominant systems for classifying books in the United States, other libraries around the world also use them. In countries that do not use a Latin alphabet, alternative systems have been created, such as the Book Classification for Chinese Libraries (BCCL) in China and the Library-Bibliographical Classification (LBC) in Russia.

Classification systems for journals also support the retrieval of journals and articles based on fields/subjects, but they are used for evaluation purposes as well. It is in this realm that classification approaches matter greatly: "reference standards obtained from questionable subject assignment might result in misleading conclusions" [27.57, p. 357]. According to *Glänzel* and *Schubert* [27.57], classifications may be produced at different levels of scholarly communication. One may take a cognitive approach, a pragmatic approach, or a scientometric approach. Both the pragmatic and scientometric approaches relate primarily to bibliometric practices, with the first related to journals and the second, individual papers. Commercial journal citation indexes currently use pragmatic subject codes for journals, and typically each subject area is also linked to observed citation patterns. Indexers who monitor these patterns may assign journals to more than one subject category or code (e. g., in Scopus, *The New England Quarterly—A Historical Review of New England Life and Letters* belongs to the history All Science Journal Classification (ASJC) code 1202 and the literature code 1208).

When comparing journals to books, classification systems like the DDC and LCC also produce a code, each approximately 6–10 digits in length. For example, the LCC number for the book titled *Uncensored War: The Media and Vietnam* is DS559.46.H35 1986. Here, the pragmatic approach to classification is also retrieval-based, but it is further subject to *literary warrant*: an LCC can only be produced on the basis of what the classified literature and controlled vocabulary of that time warrant [27.58, 59]. The first two lines, separated by a decimal, refer to the subject of the book. The third line represents the name of the author, and the last line is the book's publication date. When the library patron finds this call number in a catalog, he or she can go to a section of the library and locate the exact book.

Replicas of the same book may be on the shelf, including others related to the same topic, but the book does not appear in two different shelf locations (e. g., the history and the political science shelving areas), even if it contains information pertaining to both subjects. In sum, books differ from journals because they are normally fixed to one subject class or category.

Fast-forward to the digital age and the new Book Citation Index℠, and it is still unclear what Clarivate Analytics means by *putting books back into the library*. How will this new digital *library* contribute to an evaluation context? More specifically, how can traditional subject classification systems for books, like the DDC and LCC, support *metric* evaluations? At present, none of the traditional book classification schemes have been incorporated into the Book Citation Index℠. Instead, one finds categories and keywords, which have yet to be fully explained. For example, one can look for the book *Epicureans and Atheists in France, 1650–1729* [27.60] in a traditional library catalog, and the classification will be either an LCC B573.K67 2016 or DDC 194–dc23 (see http://lccn.loc.gov/2016008144). In the Book Citation Index℠, this same title is simply classified under the following key terms: `history`; `philosophy`; `religion`.

The level of granularity afforded by classifications such as the LCC and DDC is thus overlooked and may be problematic for book metrics, given what we know for journals. The classification of journals by field/subject is considered "one of the basic preconditions of valid scientometric analyses" [27.57, p. 357]. Journal categories are used, for example, to map the structure of science [27.61, 62], to normalize impact factor values, and to aid in the calculation of impact factor windows [27.63–65]. So far, little research has been done to reflect the role of subject classifications as a precondition for book or *biblio*-metrics [27.66–68]. This is in part due to the current structure of the Book Citation Index℠. A solution is needed, particularly for the social sciences and humanities, since these fields are more strongly represented in this index than in any other databases of the Web of Science [27.18, 66].

To circumvent the classification problem, at least two approaches have been employed. The first, devised by *Glänzel* et al. [27.67], was to match the current Web of Science classification scheme to 74 subfields from the modified Leuven-Budapest classification scheme. With this combined classification approach, clear differences were found for the citation impacts of humanities books versus those of journals in the same field. Another method involves the use of an application programming interface (API) to match titles of cited books retrieved from Scopus to the same titles recorded in the OCLC-WorldCat union catalog [27.68, 69]. Following this matching process, *Zuccala* and *White* [27.68] were thus able to classify a selection of titles from Scopus history journals (published in 1996–2000 and 2007–2011) according to their respective DDC classes.

27.4 The Indexers

While a subject classification system is essential to both the retrieval and evaluation of books, a metadata framework designed to catalog them is also needed. We separate the indexers from the subject classifiers because the decisions that these specialists make with regard to metadata also affect the practice of *biblio*-metrics, but in a different way. In short, the data that can be analyzed is only as good as how accurately it has been indexed; hence the process of indexing thousands of books in the Book Citation Index℠ has become both a research topic and a subject for serious scrutiny [27.18, 70, 71].

Not long after the Book Citation Index℠ was launched, *Gorraiz* et al. [27.18] performed some test analyses and found that "out of the almost 30 000 *books* retrieved [for] the publication period 2005–2011, only about 1100 provide[d] author affiliations" [27.18, pp. 1390–1392]. In addition to missing address information, the researchers noticed that the term `book` as a registered document type had the potential to be confusing, especially if edited books were not carefully delineated from the whole book content of monographs. In cases where there was no clear delineation, there was further potential for false interpretations: for example,

> selecting *books* as document type and sorting the results by most cited... present[ed] a list of the most cited books as a whole, but disregard[ed] all the citations to single chapters.... Similarly, sorting *book chapters* by times cited omit[ted] whole-book citations. [27.18, p. 1392]

Index-focused research also suggests that there can be a problem with underrating or overrating the citation impact monographs if individual chapters from specific monographs are counted separately [27.70].

In comparison to journal articles, monographs are difficult to index because they typically belong to bibliographic families [27.71]. Unlike journal articles, they can be revised and reprinted as new editions. In the past, many book catalogs have benefited from guidelines such as the Functional Requirements for Bibliographic Records (FRBR) [27.72, 73]; hence *Zuccala* et al. [27.71] suggest that the Book Citation Index℠ can benefit as well. With *Tillet*'s [27.72, 73] conception of the FRBR model, every monograph in a bibliographic family is a physical entity or *manifestation*, with its own International Standard Book Number (ISBN). If several different *manifestations* share the same intellectual properties, they are *expressions* (editions), and together all derivative *expressions* (editions) relate to one *work* [27.71]. A *work* is therefore the progenitor for a bibliographic family—the starting point for all ideational and semantic content [27.74]. Any new *expression*, or edition, of a monograph that deviates significantly from the progenitor is called a new *work*.

The Book Citation Index℠ might potentially be revised to follow FRBR, so that every *expression*, or edition, of a monograph is indexed according to its full set of manifestations (i. e., all ISBNs per physical type), its own unique *expression* identifier, and its shared *work* identifier. For each manifestation of a particular book there is, however, a specific problem to consider. Books, unlike journal articles, do not have their own unique Digital Object Indicators (DOIs). Currently, the ISBN is the most frequently used identifier for retrieving and matching identical book titles recorded in different databases [27.69, 75]. It is important, however, to recognize that (a) ISBNs do need to be registered, as this gives a clear idea of how many times a book has been reprinted, and (b) publication and citation counts should not be calculated at the level of the ISBN, as this does not correspond to the intellectual content of a work, but rather its physical container. With a proposed FRBR-guided version of the Book Citation Index℠, all ISBNs per book would be present, but the addition of new identifiers means that bibliometricians might have more accurate options for counting books at either the *expression* level or the *work* level. *Zuccala* et al. explain why this matters [27.76, p. 156]:

> The value in calculating indicators at different bibliographic levels is that it can help to identify whether or not a specific expression or edition of a monograph is receiving more attention than the work as a whole. For instance, one specific expression of a work may be cataloged in libraries, used, referred to, or reviewed more frequently than another. This could be the literal translation of a non-English edition of a work to English, with the new English-language edition potentially having a wider appeal. For some types of translated works, in fact, an author might even have more than one metric profile.

27.5 The Indicator Constructionists

Indicator constructionists are researchers who develop indicators for use in quantitative research evaluation systems. This group of experts differs from the monitors because they are less intent on describing approaches to *biblio*-metrics and more committed to identifying and promoting real methodological solutions. Progress in this regard has been greatly aided by technological advancements and the emergence of new data sources, for example, the Book Citation Index℠, the Scopus index of books, Google Books, Google Scholar, OCLC-WorldCat, Goodreads, Amazon Reviews, and national academic repositories [27.68, 77–79].

The process of evaluating books depends, however, on more than just data. When a particular data source is used to advance an indicator, advocates of that indicator need to reflect to some degree on a theory [27.80, 81]. According to *Zuccala* [27.81], the main task of the humanistic *biblio*-metrician, or book evaluation specialist, is not to simply "expand his/her metric toolkit, but to first examine the term *indicator*" [27.81, p. 159]. *Gingras* [27.80] upholds this notion by explaining that if an indicator serves as a proxy for a concept, it must be closely aligned with the concept or object that it is designed to measure. The primary, ongoing difficulty is that "the reality behind the concept [might] change over time and/or place" [27.80, p. 113]). In *Van der Weel*'s *Changing out textual minds*, we are reminded of this fact for books [27.4, p. 2]:

> Digitisation of textual transmission is proceeding so rapidly that already the consequences are huge and all-encompassing, indeed revolutionary. As reading practices move on line the once discrete products of the print world all become part of the digital textual *docuverse*, and that docuverse in turn becomes part of the all-digital array of mediums converged on the WorldWide Web.

Will bibliometric evaluations manage to keep up with this revolution?

In terms of data and theory, the research community thus far has taken two paths toward developing book indicators. One route has been to focus on the traditional citation—e. g., extracting citations to books as non-sourced items in commercial indexes [27.43, 82, 83]. The other has been to avoid the citation and focus on book reviews [27.78, 84], publisher quality, and specialization [27.85, 86], and library holding counts [27.87, 88].

27.5.1 Citations

In principle, a new data source like the Book Citation Index℠ could seem like the perfect solution for developing indicators for books. Still, there are certain factors to take into account. Research has shown that citation patterns for books differ from those of journal articles [27.15, 85, 89], and that in comparison to journal articles, the citation age for books is longer [27.67]. The role that a book plays within a particular scholarly communication system also differs depending on the discipline under study [27.50]. And even within different disciplines, there can be citation effects related to book types [27.90, 91], language and internationalization [27.92], and variations in authorship patterns [27.93].

With the Book Citation Index℠, the drawbacks to developing new indicators rest with the selection bias of monographs published in the English language, a high concentration of books printed by large publishers, and unclear distinctions between different editions and translations of the same monograph [27.76, 85]. There is, however, at least one benefit to this index, in that it enables large-scale comparative analyses of citation distributions for both monographs and journal articles [27.67, 94].

In 2004, Google launched two revolutionary services: Google Books and Google Scholar. Both services not only offer quick and easy access to scientific literature, but also give researchers an opportunity to engage in full-text searching. This in turn enhances the ability to capture citations from a great variety of research sources. The downside to these platforms is that mechanisms by which researchers can identify citations often produce false positives and prevent opportunities for large-scale analyses [27.50]. Thus, when using Google Books or Google Scholar, researchers suggest that it may be wise to use citation data only as a complement to peer review [27.95].

Citations have been used outside the scholarly communication system to assess the non-scientific impact of research where scholarship may be targeting a non-scholarly public, intentionally or not. For instance, citations from Wikipedia, which are now part of the set of indicators offered by the platform Altmetric.com, have been suggested as a means to capture extra evidence of impact [27.96]. The issue of scarce counts, however, makes the *Wiki-cite* unreliable for use in a real research assessment exercise. There are also many syllabuses and teaching materials that include citations to research, which means that books may be further measured in terms of their educational impact [27.79]. Since correlations between educational-based and research-based citations tend to be low, educational impact is arguably a different type of measure, warranting further investigation on its own.

27.5.2 Publisher Prestige or Quality

With the study of books, an analogy may be drawn between journals and publishers. However, unlike measures for journals—i. e., the journal impact factor (JIF) [27.97], the source normalized impact per paper (SNIP) [27.98], and the SCImago Journal Rank (SJR) [27.99]—there is currently no similar impact-based quantitative indicator for books. The main focus, therefore, has been to assess publisher prestige or quality instead of impact, and to direct this toward expert (scholars') opinions rather than citations [27.100, 101]. Proponents of this research area argue that citation data does not accurately capture the impact of books, and that this is particularly the case in many humanities disciplines, where the goal is not to create impact per se, but to influence further academic thinking and/or debate [27.102]. The expert-oriented approach is or has largely been inspired by the work of *Nederhof* et al. [27.103], who first studied publisher quality within the field of linguistics. In this study, scholars from the Netherlands, Flanders, and worldwide were invited to participate in a survey. With the results, *Nederhof* et al. [27.103] were able to differentiate among the three populations and gain insight into the locality of prestige, language biases, and disciplinary differences—all issues considered to be highly relevant within the social sciences and humanities [27.44].

As a result, we have seen at least one indicator that has been developed and proposed for the evaluation of book publisher *quality* and *prestige*. In the research by *Giménez-Toledo* et al. [27.101], 14 questions were sent to various academics/scientists from different research fields as part of a survey that was structured in three blocks:

1. Profile of the respondent
2. Evaluation of the quality of a publisher with scientific publications
3. Evaluation of the publishing process of a publishing house with scientific publications [27.101, p. 67].

Following the survey, the data were used to calculate what the authors term an "*Indicator of Quality* *for Publishers according to Experts (ICEE)*" [27.101, p. 68].

Not all scholars agree with the focus on publishers in the development of book metrics. *Verleysen* and *Engels* [27.104] indicate, for instance, that publishing houses are not the most suitable level of aggregation, and argue that it is impractical to perform a *quality analysis* for each and every book title after it has been published. As a compromise, they suggest creating a label for peer-reviewed monographs so as to ensure that researchers and evaluators know that a certain level of formal quality has been ascertained prior to publication. In this way, greater emphasis is placed on the precondition for book quality rather than a metric analysis of quality ex post facto—a point which continues to be under international discussion [27.100].

Yet another area of interest has been the study of publisher specialization [27.101, 105, 106]. To understand specialization, one approach has been to take a specific unit of analysis, such as the book chapter, and develop mapping techniques designed to visualize their disciplinary profiles [27.107]. Network maps, which follow directed citations to books from journals, have also been used to identify the specialization of both commercial and university presses [27.86]. In the research of *Mañana-Rodríguez* and *Gimenez-Toledo* [27.105], the tension between publisher specialization and multidisciplinarity has been measured using what the authors call an "entropy-based indicator" [27.105, p. 19]. When a publisher publishes books in fewer fields, its specialization increases, whereas its multidisciplinary profile may also increase if there is unevenness in its distribution of titles across different fields.

To date, only a few publisher rankings have been produced, and only for certain research fields [27.86, 108]. Within a specific time frame, a ranking of publishers may be calculated on the basis of their overall citations received or average citations per book [27.86]. However, if a ranking is based on citations, typically the most powerful English-language publishing houses are listed. This is because a large majority of publishing houses tend to have high rates of un-citedness [27.89]. Citations only reveal a small portion of what is happening in the publishing industry. A careful ranking procedure must therefore consider the fact that every publishing house or press differs in terms of economic capital, symbolic capital, and geographical reach [27.2, 86].

27.5.3 Book Reviews

According to *Lindholm-Romantschuk* [27.30], the difficulty attached to finding an appropriate quantitative indicator for assessing book quality is that processes of formal assessment are and already have been taking place. For many years, "the evaluation of scholarly monographs [has been] contained within the system of academic reviewing" [27.30, p. 36]. Book reviews still play an important role in the reception of scholarly monographs, but the lack of esteem attached to reviews has sometimes led to legitimate concerns regarding their judicious value [27.20]. It can be useful, therefore, to filter out specific types of reviews by focusing on those that are more *scholarly*, or at least those that researchers agree upon as having familiar or trusted scholarly characteristics [27.28, 39]. Evidence of scholarliness can be assessed, for example, by the degree to which a reviewer includes references to other academic sources in addition to the book under review [27.84].

Yet another way to import book reviews into an evaluation context is to make use of them as *mega-citations*. *Zuccala* et al. [27.109] have introduced a theory of *mega-citation* which explains how book reviews may be transformed into quantitative indicators based on a full-text analysis of reviewer comments. One drawback to working with *mega-citations* is that full-text reviews published in journals are largely inaccessible in electronic form. In light of this problem, some bibliometricians have found that public and socially motivated book reviews are a better option, particularly those published on sites such as Amazon.com or Goodreads. Public reviews are especially useful for indicating the degree to which a scholarly book has become visible online and has become a topic for social engagement [27.75, 110]. Both scholarly and public reviews can always be used in conjunction with other types of indicators (e.g., publisher quality and/or citations), but for improving the coverage of books in commercial citation indexes, preference is given to the scholarly review [27.111].

27.5.4 Library Holding Counts

To date, the most promising of all book-based indicators is the library holding count [27.71, 87, 88], which *White* et al. refer to as the *libcitation* [27.88, p. 1083]. A theory of libcitation rationalizes that a holding count or set of holding counts in library catalogs might be used to indicate and calculate the perceived cultural benefit of a book or books. The advantage of this measure is that it "can make an author in the humanities look good", particularly if (s)he is not well represented in other types of databases such as Web of Science, Google Scholar, or Scopus. *White* further explains that [27.88, p. 1084]:

> On the book front, libcitations reflect what librarians know about the prestige of publishers,

the opinions of reviewers, and the reputations of authors. The latter may be colored by, for example, authors' academic affiliations, previous sales, prizes, awards, distinguished appointments, mass media coverage, Web presence, and citedness. All of these are signals of what readers are likely to want, and librarians must be attuned to them.

When working with this indicator, at least two different methodological approaches are possible. *Torres-Salinas* and *Moed* [27.87] focus on library holding counts at the publisher level, while *White* et al. [27.88] propose developing it at the book level. At the book level, holding counts have much more power to discriminate between books than citation counts. Records for books tend to be more plentiful in library catalogs than in citation indexes, particularly in a union catalog like OCLC-WorldCat [27.112]. Libcitation counts for individual scholars or academic departments can be field-normalized or assigned to percentiles just as citations are. By determining how many libcitations a book needs in order to reach a 90th- or 50th-percentile cutpoint in its main Dewey class, one can observe its cultural impact, or degree of fame relative to other titles from the same class [27.71]. Research also points to the fact that libcitations and citations can be statistically correlated, but one is likely to find a weak, albeit significant result [27.68]. Both the citation and the libcitation capture a certain amount of scholarly impact in common, but this seems to be truer when holding counts are obtained from academic libraries rather than other types of libraries. Another study using the PlumX suite of altmetrics, now shows that out of 18 types of indicators for books, including citations, downloads, views, and social media mentions, the most informative is the library holding count [27.113].

27.6 Integrating Book Metrics into Evaluation Practices

For some time, the social sciences and humanities have been either only partially assessed or neglected entirely due to the lack of data available for developing promising book metrics. Acceptance of this fact grew in part because of the increasing value of journal articles (in most fields), notwithstanding the long tradition of relying on journal citation indexes for many international research evaluation procedures. Fortunately, this did not stop some of the early bibliometric monitors from examining the role of books in book-oriented research disciplines, nor did it prevent commercial organizations like Elsevier and Clarivate Analytics (formerly Thomson Reuters) from addressing the data gap by developing a Scopus index of books and Book Citation Index℠. Subject classifiers and indexers now have ample reason to step to the forefront, not only to apply research to these indexes, but to lead the bibliometrics community forward to an improved situation, one in which the metric exploration of books and their publishers is no longer an aspiration, but an established reality.

Still, the integration of books into evaluation practices will never be left solely to commercial data providers or researchers. National policymakers are stakeholders in the evaluation game and also play a role. Research by *Giménez-Toledo* et al. [27.77] and *Williams* et al. [27.8] provides valuable overviews of how countries across Europe have recently been implementing policies and strategies for book-based evaluations. In the United Kingdom, most scholarly books are submitted to panels C (social sciences) and D (humanities) of the panel-based Research Excellence Framework (REF). Since the panels (as well as sub-panels) take into consideration what is most valued in these broader disciplines, a qualitative approach to evaluation is used. A different approach is taken in Spain, Denmark, and Finland, where evaluation procedures for books are based on league tables or *authority* lists of publishers. Panels of experts are recruited here as well, but are invited to participate in the development of such lists. The publisher lists are then used to benchmark the value of a monograph submitted to each country's respective evaluation exercise. Other countries like Flanders (Belgium) implement a point system by which a book's value is weighted (e. g., monographs receive four points, while edited books receive one point). Norway uses a mixed-methods approach, where publishers and journals are divided into two levels, where a level 2 designation is the most selective. Depending on the level, a monograph will receive either five points (i. e., for a book with a level 1 publisher) or eight points (i. e., for a book with a level 2 publisher). Denmark generally follows Norway's approach; hence, with a similar system, a fraction of funding each year is allotted to Danish university departments that achieve the highest numbers of points.

More often than not, these evaluation policies are designed for practical purposes. Again, it is simply impractical to assess the individual contribution, quality, influence, or impact of every monograph at a national or international level. This issue, together with the uncertainty of applying citation analysis to books, and criticisms coming from social scientists and humanities researchers, has prevented the widespread development and use of citation-based indicators. Policymakers have thus been keen to disregard the citation, including many other practices, in favor of focusing on publisher status (i. e., as per the league, or *authority* tables). As a result, certain challenges related to book metrics have yet to be addressed. We are presently at the stage where disparities in data coverage [27.113] and low correlations between citations to books and alternative indicators of their impact [27.68, 96, 110] remain difficult to interpret. With citation indicators alone, differences per database at least show moderately significant correlations [27.95].

From a research perspective, it is clear, then, that more work is needed to improve upon the subject classification of books, both in commercial and in national indexes, and to ensure that record keeping is complete (e. g., indexes that include author affiliations and show how books belong to bibliographic families). Scholars who work with these indexes—i. e., the indicator constructionists—are urged to remain steadfast at uncovering, refining, and emphasizing different elements related to the influence or impact of books. Their biggest challenge, however, may not be technical or data-oriented, but cultural, in nature.

Citation-based indicators have long been associated with research assessment schemes directed toward the natural and exact sciences. Journal articles and their citations received accommodate research communities grounded upon previous work and rapid progress: a Kuhnian model of normal science. By contrast, books and their reviews fit within a *social* view of scholarship. Here the standards are based on the perceptions of peers; it is the academic peer who determines the value of a work. In theoretical disciplines, where books are most prominent, this community-based reflexivity, inherent in the overall reflexive nature of the social sciences and humanities, is likely to remain a primary strength [27.114]. When bibliometric approaches to evaluation focus on complementing this strength, and also recognizing a book's broader (i. e., educational, social, literary) influence or impact, book-based scholarship will evolve not in response to perceived faults in the evaluation system, but because different aspects of the truth will become more evident.

References

27.1 E.L. Eisenstein: *The Printing Press as Agent of Change* (Cambridge Univ. Press, Cambridge 2013), https://www-cambridge-org.ep.fjernadgang.kb.dk/core/books/the-printing-press-as-an-agent-of-change/7DC19878AB937940DE13075FE839BDBA

27.2 J.B. Thomson: *Books in the Digital Age: The Transformation of Academic and Higher Education Publishing in Britain and the United States* (Polity, Cambridge 2005)

27.3 M. Mann: *Introduction to Cataloging and the Classification of Books*, 2nd edn. (American Library Association, Chicago 1930)

27.4 A. Van der Weel: *Changing our Textual Minds, Towards a Digital Order of Knowledge* (Manchester Univ. Press, Manchester 2011)

27.5 E. Garfield: Is information retrieval in the arts and humanities inherently different from that in science? The effect that ISI®'s Citation Index for the arts and humanities is expected to have on future scholarship, Libr. Q. **50**(1), 40–57 (1979)

27.6 R.A. Day: *How to Write and Publish a Scientific Paper*, 3rd edn. (Cambridge Univ. Press, Cambridge 1989)

27.7 L.A. Coser: Publishers as gatekeepers of ideas, Ann. Am. Acad. Polit. Soc. Sci. **421**(1), 14–22 (1975)

27.8 G. Williams, A. Basso, I. Galleron, T. Lippiello: More, less or better: The problem of evaluating books in SSH research. In: *The Evaluation of Research in Social Sciences and Humanities. Lessons from the Italian Experience*, ed. by A. Bonaccorsi (Springer, Cham 2018) pp. 133–158

27.9 C. Basili, L. Lanzillo: Research quality criteria in the evaluation of books. In: *The Evaluation of Research in Social Sciences and Humanities. Lessons from the Italian Experience*, ed. by A. Bonaccorsi (Springer, Cham 2018) pp. 159–184

27.10 S. Chodorow: The Once and Future Monograph, Scholarly Monograph in Crisis or How Can I Get Tenure if You Won't Publish My Book?, http://www.arl.org/resources/pubs/specscholmono/Chodorow.shtml (1997), Paper presented at the Association of Research Libraries, Washington, DC

27.11 S. Fish: *Is there a Text in this Class? The Authority of Interpretive Communities* (Harvard Univ. Press, Harvard 1980)

27.12 J. Fry, C. Oppenheim, C. Creaser, W. Johnson, M. Summers, S. White, G. Butters, J. Craven, J. Griffiths, D. Hartley: *Communicating knowledge: How and Why Researchers Publish and Disseminate Their Findings* (Research Information Network, London 2009)

27.13 L. Butler, M.S. Visser: Extending citation analysis to non-source items, Scientometrics **66**(2), 327–343 (2007)

27.14 B. Cronin, K. La Barre: Mickey Mouse and Milton: Book publishing in the humanities, Learn. Publ. **17**(2), 85–98 (2004)

27.15 B. Cronin, H. Snyder, H. Atkins: Comparative citation rankings of authors in monographic and

journal literature: A study of sociology, J. Doc. **53**(3), 263–273 (1997)
27.16 L. Taycher: Books of the world, stand up and be counted! All 129,864,880 of you. http://booksearch.blogspot.dk/2010/08/books-of-world-stand-up-and-be-counted.html (2010)
27.17 Clarivate Analytics: Putting Books Back into the Library: Completing the Research Picture. The Book Citation Index℠, http://wokinfo.com/products_tools/multidisciplinary/bookcitationindex/ (2017)
27.18 J. Gorraiz, P. Purnell, W. Glänzel: Opportunities and limitations of the book citation index, J. Am. Soc. Inf. Sci. Technol. **64**(7), 1388–1398 (2013)
27.19 H.F. Moed: The future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review, Sci. Public Policy **34**(8), 575–583 (2007)
27.20 N. Glenn: On the misuse of book reviews, Contemp. Sociol. **7**(3), 254–255 (1978)
27.21 A. Natowitz, P.C. Wheeler: Evaluating review content for book selection: An analysis of American history reviews in Choice, American Historical Review, and Journal of American History, Coll. Res. Libr. **58**(4), 322–335 (1997)
27.22 J.M. Parker: Scholarly book reviews in literature journals as collection development sources for librarians, Collect. Manag. **11**(12), 41–57 (1989)
27.23 J. Serebnick: Selection and holdings of small publishers' books in OCLC libraries: A study of the influence of reviews, publishers and vendors, Libr. Q. **62**(3), 259–294 (1992)
27.24 J.W. Cortada: Five ways to be a terrible book reviewer, J. Sch. Publ. **30**(1), 34–37 (1998)
27.25 V.E. Hargrave: A comparison of reviews of books in the social sciences in general and scholarly periodicals, Libr. Q. **18**, 206–217 (1948)
27.26 D. Motta-Roth: Discourse analysis and academic book reviews: A study of text and disciplinary cultures. In: *Genre Studies in English for Academic Purposes*, ed. by I. Fortanet (Universitat Jaume, Castelló de la Plana 1998) pp. 29–58
27.27 H. Regnery: Bias in book reviewing and book selection, Am. Libr. Assoc. Bull. **60**, 57–62 (1966)
27.28 J. Hartley: Reading and writing book reviews across disciplines, J. Am. Soc. Inf. Sci. Technol. **57**(9), 1194–1207 (2006)
27.29 A. Spink, D. Robbins, L. Schamber: Use of scholarly book reviews: implications for electronic publishing and scholarly communication, J. Am. Soc. Inf. Sci. Technol. **49**(4), 364–374 (1998)
27.30 Y. Lindholm-Romantschuk: *Scholarly Book Reviewing in the Social Sciences and Humanities. The Flow of Ideas Within and Among Disciplines* (Greenwood, Westport 1998)
27.31 T.D. Bilhartz: In 500 words or less: Academic book reviewing in American history, Hist. Teach. **17**(4), 525–536 (1984)
27.32 D.T. Champion, M.F. Morris: A content analysis of book reviews in the AJS, ASR, and social forces, Am. J. Sociol. **78**(5), 1256–1265 (1973)
27.33 L.E. Riley, E.A. Spreitzer: Book reviewing in the social sciences, Am. Sociol. **5**(4), 358–363 (1970)
27.34 W.E. Snizek, E.R. Fuhrman: The evaluative content of book reviews in the American Journal of Sociology, Contemporary Sociology, and Social Forces, Contemp. Sociol. **8**(3), 339–340 (1979)
27.35 W.E. Snizek, E.R. Fuhrman: Some factors affecting the evaluative content of book reviews in sociology, Am. Sociol. **14**(2), 108–114 (1979)
27.36 E. Garfield: Citation indexes for retrieval and research evaluation. In: *Consensus Conference on the Theory and Practice of Research Assessment* (1996), http://www.garfield.library.upenn.edu/papers/ciretreseval-capri.html
27.37 V. Diodato: Impact and scholarliness in arts and humanities book reviews: A citation analysis. In: *Proc. 47th Annu. Meet. Am. Soc. Inf. Sci.*, Vol. 21 (1984) pp. 217–221
27.38 A.J. Nederhof: Bibliometric monitoring of research performance in the social sciences and the humanities: A review, Scientometrics **66**(1), 81–100 (2006)
27.39 J. Nicolaisen: The scholarliness of published peer reviews: A bibliometric study of book reviews in selected social science fields, Res. Eval. **11**(3), 129–140 (2002)
27.40 A. Schubert, S. Zsindely, A. Telcs, T. Braun: Quantitative analysis of a visible tip of the iceberg: Book reviews in chemistry, Scientometrics **6**(6), 433–443 (1984)
27.41 J.W. Thompson: The death of the scholarly monograph in the humanities? Citation patterns in literary scholarship, Libri **52**, 121–136 (2002)
27.42 G. Lewison: Evaluation of books as research output in the history of medicine, Res. Eval. **10**(2), 89–95 (2001)
27.43 B. Hammarfelt: Interdisciplinarity and the intellectual base of literature studies: Citation analysis of highly cited monographs, Scientometrics **86**(3), 705–725 (2011)
27.44 D. Hicks: The four literatures of social science. In: *Handbook of Quantitative Science and Technology Research*, ed. by H.F. Moed, W. Glänzel, U. Schmoch (Springer, Dordrecht 2004) pp. 473–496
27.45 A.V. Rubio: Scientific production of Spanish universities in the fields of social sciences and language, Scientometrics **24**(1), 3–19 (1992)
27.46 P. Williams, I. Stevenson, D. Nicholas, A. Watkinson, I. Rowlands: The role and future of the monograph in arts and humanities research, ASLIB Proc. New Inf. Perspect. **61**(1), 67–82 (2009)
27.47 J. Guillory: Evaluating scholarship in the humanities: Principles and procedures, ADE Bulletin **137**, 18–33 (2005)
27.48 E. Archambault, E. Vignola Gagné: *The use of Bibliometrics in the Social Sciences and Humanities, Science-Metrix, Final Report* (Social Sciences and Humanities Research Council of Canada (SSHRCC), Montreal 2004)
27.49 M. Huang, Y. Chang: Characteristics of research output in the social sciences and humanities: from a research evaluation perspective, J. Am. Soc. Inf. Sci. Technol. **59**(11), 1819–1828 (2008)

27.50 K. Kousha, M. Thelwall: Google book citation for assessing invisible impact?, J. Am. Soc. Inf. Sci. Technol. **60**(8), 1537–1549 (2009)
27.51 D. Hicks, J. Wang: Towards a bibliometric database for the social sciences and humanities, http://works.bepress.com/cgi/viewcontent.cgi?article=1017context=diana_hicks (2009)
27.52 V. Larivière, B. Macaluso: Improving the coverage of social science and humanities researchers' output: The case of the Érudit journal platform, J. Am. Soc. Inf. Sci. Technol. **62**(12), 2437–2442 (2011)
27.53 H.F. Moed, J. Linmans, A. Nederhof, A. Zuccala, C. Lopez Illescas, F. de Moya Anegon: *Options for a Comprehensive Database of Research Outputs in the Social Sciences and Humanities, Research Report to the Project Board of the Scoping Study "Towards a Bibliometric Database for the Social Sciences and Humanities"* (Standing Committees for the Social Sciences and the Humanities of the European Science Foundation (ESF), Strasbourg 2009)
27.54 G. Sivertsen, B. Larsen: Comprehensive bibliographic coverage of the social sciences and humanities in a citation index: An empirical analysis of the potential, Scientometrics **91**(2), 567–575 (2012)
27.55 E. Garfield: Citation indexes for science: A new dimension in documentation through association of ideas, Science **122**(3159), 108–111 (1955)
27.56 OCLC Online Computer Library Center, Inc.: Map Dewey (DDC) Classification to LC (LCC), http://www.questionpoint.org/crs/html/help/pl/ask/ask_map_ddctolcc.html (2017)
27.57 W. Glänzel, A. Schubert: A new classification scheme of science fields and subfields designed for scientometric evaluation purposes, Scientometrics **56**(3), 357–367 (2003)
27.58 C. Beghtol: Semantic validity: Concepts of warrant in bibliographic classification systems, Libr. Resour. Tech. Serv. **30**(2), 109–125 (1986)
27.59 L.M. Chan: *A Guide to the Library of Congress Classification*, 5th edn. (CO: Libraries Unlimited, Englewood 1999)
27.60 A.C. Kors: *Epicureans and Atheists in France* (Cambridge Univ. Press, Cambridge 2016) pp. 1650–1729
27.61 K.W. Boyack, R. Klavans, K. Börner: Mapping the backbone of science, Scientometrics **64**(3), 351–374 (2005)
27.62 L. Leydesdorff, I. Rafols: A global map of science based on the ISI subject categories, J. Am. Soc. Inf. Sci. Technol. **60**(2), 348–362 (2009)
27.63 P. Dorta González, M.I. Dorta González: Comparing journals from different fields of science and social science through a JCR subject categories normalized impact factor, Scientometrics **95**(2), 645–672 (2013)
27.64 M.I. Dorta González, P. Dorta González: Do fixed citation windows affect the impact maturation rates of scientific journals?, Invest. Bib. **30**(68), 73–89 (2015)
27.65 W. Glänzel, H. Moed: Journal impact measures in bibliometric research, Scientometrics **53**(2), 171–193 (2002)
27.66 P.-S. Chi, W. Jueris, B. Thijs, W. Glanzel: Book bibliometrics – A new perspective and challenge in indicator building based on the Book Citation Index. In: *Proc. ISSI 2015 Istanbul: 15th Int. Soc. Sci. Informetr. Conf*, ed. by A.A. Salah, Y. Tonta, A.A. Akdag Salah, C.R. Sugimoto, U. Al (Bogaziçi Univ. Printhouse, Istanbul 2015) pp. 1161–1169
27.67 W. Glänzel, B. Thijs, P.-S. Chi: The challenges to expand bibliometric studies from periodical literature to monographic literature with a new data source: The Book Citation Index, Scientometrics **109**(21), 65–2179 (2016)
27.68 A. Zuccala, H.D. White: Correlating libcitations and citations in the humanities with WorldCat.org and scopus data. In: *Proc. 15th Int. Soc. Sci. Informetr. (ISSI)*, ed. by A.A. Salah, Y. Tonta, A.A. Akdag Salah, C. Sugimoto, U. Al (Bogazici Univ., Istanbul 2015) pp. 305–316
27.69 A. Zuccala, R. Cornacchia: Data matching, integration, and interoperability for a metric assessment of monographs, Scientometrics **108**(1), 465–484 (2016)
27.70 L. Leydesdorff, U. Felt: "Books" and "book chapters" in the book citation index (BKCI) and science citation index (SCI, SoSCI, AHCI), Proc. Am. Soc. Inf. Sci. Technol. **49**(1), 1–7 (2012)
27.71 H. White, A. Zuccala: Libcitations, WorldCat, cultural impact, and fame, J. Assoc. Inf. Sci. Technol. (2018), https://doi.org/10.1002/asi.24064
27.72 B. Tillet: Bibliographic relationships. In: *Relationships in the Organization of Knowledge*, ed. by C.A. Bean, R. Green (Kluwer, Dordrecht 2001) pp. 19–35
27.73 B. Tillet: What is FRBR? A conceptual model for the bibliographic universe, Aust. Libr. J. **54**(1), 24–30 (2005)
27.74 R.P. Smiraglia: *The Nature of a Work: Implications for the Organization of Knowledge* (Scarecrow, London 2001)
27.75 K. Kousha, M. Thelwall, S. Abdoli: Goodreads reviews to assess the wider impacts of books, J. Assoc. Inf. Sci. Technol. **68**(8), 2004–2016 (2017)
27.76 A. Zuccala, M. Breum, K. Bruun, B.T. Wunsch: Metric assessments of books as families of works, J. Assoc. Inf. Sci. Technol. **69**(1), 146–157 (2018)
27.77 E. Giménez-Toledo, J. Mañana-Rodríguez, T.C.E. Engels, P. Ingwersen, J. Polonen, G. Sivertsen, F.T. Verleysen, A.A. Zuccala: Taking scholarly books into account. Current developments in five European countries, Scientometrics **107**(2), 685–699 (2016)
27.78 K. Kousha, M. Thelwall: Can Amazon.com reviews help to assess the wider impacts of books?, J. Assoc. Inf. Sci. Technol. **67**(3), 566–581 (2016)
27.79 K. Kousha, M. Thelwall: An automatic method for assessing the teaching impact of books from online academic syllabi, J. Assoc. Inf. Sci. Technol. **67**(12), 2993–3007 (2016)

27.80 Y. Gingras: Criteria for evaluating indicators. In: *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*, ed. by B. Cronin, C.R. Sugimoto (MIT Press, Cambridge 2014) pp. 109–125

27.81 A. Zuccala: Inciting the metric oriented humanist: Teaching bibliometrics in a faculty of humanities, Educ. Inf. **32**(2), 149–164 (2016)

27.82 L. Butler, M.S. Visser: Extending citation analysis to non-source items, Scientometrics **66**(2), 327–343 (2012)

27.83 A.J.M. Linmans: Why with bibliometrics the Humanities does not need to be the weakest link. Indicators for research evaluation based on citations, library holdings, and productivity measures, Scientometrics **83**(2), 337–354 (2010)

27.84 A. Zuccala, T. van Leeuwen: Book reviews in humanities research evaluations, J. Assoc. Inf. Sci. Technol. **62**(10), 1979–1991 (2011)

27.85 D. Torres-Salinas, N. Robinson-Garcia, J.M. Campanario, E.D. López-Cózar: Coverage, field specialisation and the impact of scientific publishers indexed in the Book Citation Index, Online Inf. Rev. **38**(1), 24–42 (2014)

27.86 A. Zuccala, R. Guns, R. Cornacchia, R. Bod: Can we rank scholarly book publishers? A bibliometric experiment with the field of history, J. Assoc. Inf. Sci. Technol. **66**(7), 1333–1347 (2014)

27.87 D. Torres-Salinas, H.F. Moed: Library catalog analysis as a tool in studies of social sciences and humanities: An exploratory study of published book titles in economics, J. Informetr. **3**(1), 9–26 (2009)

27.88 H.D. White, S.K. Boell, H. Yu, M. Davis, C.S. Wilson, F.T.H. Cole: Libcitations: A measure for comparative assessment of book publications in the humanities and social sciences, J. Am. Soc. Inf. Sci. Technol. **60**(6), 1083–1096 (2009)

27.89 D. Torres-Salinas, N. Robinson-García, A. Cabezas-Clavijo, E. Jiménez-Contreras: Analyzing the citation characteristics of books: Edited books, book series and publisher types in the book citation index, Scientometrics **98**(3), 2113–2127 (2014)

27.90 S. Milojević, C.R. Sugimoto, V. Larivière, M. Thelwall, Y. Ding: The role of handbooks in knowledge creation and diffusion: A case of science and technology studies, J. Informetr. **8**(3), 693–709 (2014)

27.91 M. Thelwall, P. Sud: No citation advantage for monograph-based collaborations?, J. Informetr. **8**(1), 276–283 (2014)

27.92 T.C.E. Engels, T.L.B. Ossenblok, E.H.J. Spruyt: Changing publication patterns in the social sciences and humanities, Scientometrics **93**(2), 373–390 (2012)

27.93 T.L.B. Ossenblok, F.T. Verleysen, T.C.E. Engels: Coauthorship of journal articles and book chapters in the social sciences and humanities, J. Assoc. Inf. Sci. Technol. **65**(5), 882–897 (2014)

27.94 P.-S. Chi: Differing disciplinary citation concentration patterns of book and journal literature?, J. Informetr. **10**(3), 814–829 (2016)

27.95 K. Kousha, M. Thelwall, S. Rezaie: Assessing the citation impact of books: The role of Google Books, Google Scholar, and Scopus, J. Am. Soc. Inf. Sci. Technol. **62**(11), 2147–2164 (2011)

27.96 K. Kousha, M. Thelwall: Are Wikipedia citations important evidence of the impact of scholarly articles and books?, J. Assoc. Inf. Sci. Technol. **68**(3), 762–779 (2017)

27.97 E. Garfield: The history and meaning of the journal impact factor, JAMA J. Am. Med. Assoc. **295**(1), 90–95 (2006)

27.98 H.F. Moed: Measuring contextual citation impact of scientific journals, J. Informetr. **4**(3), 265–277 (2010)

27.99 B. González-Pereira, V.P. Guerrero-Boteb, F. Moya-Anegónc: A new approach to the metric of journals' scientific prestige: The SJR indicator, J. Informetr. **4**(3), 379–391 (2010)

27.100 E. Giménez-Toledo, A. Román-Román: Assessment of humanities and social sciences monographs through their publishers: A review and a study towards a model of evaluation, Res. Eval. **18**(3), 201–213 (2009)

27.101 E. Giménez-Toledo, C. Tejada-Artigas, J. Mañana-Rodríguez: Evaluation of scientific books' publishers in social sciences and humanities: Results of a survey, Res. Eval. **22**(1), 64–77 (2012)

27.102 A. Zuccala: Quality and influence in literary work: evaluating the 'educated imagination', Res. Eval. **21**(3), 229–241 (2012)

27.103 A. Nederhof, M. Luwel, H. Moed: Assessing the quality of scholarly journals in linguistics: An alternative to citation-based journal impact factors, Scientometrics **51**(1), 241–265 (2001)

27.104 F.T. Verleysen, T.C.E. Engels: A label for peer-reviewed books, J. Assoc. Inf. Sci. Technol. **64**(2), 428–430 (2013)

27.105 J. Mañana-Rodríguez, E. Giménez-Toledo: Specialization and multidisciplinarity of scholarly book publishers: differences between Spanish University Presses and other scholarly publishers, Scientometrics **114**(1), 19–30 (2018)

27.106 E. Giménez-Toledo, J. Mañana-Rodríguez, C.-M. Tejada-Artigas: Scholarly publishers indicators: Prestige, specialization and peer review of scholarly book publishers, Prof. Inf. **24**(6), 855–860 (2015)

27.107 D. Torres-Salinas, R. Rodríguez-Sánchez, N. Robinson-García, J. Fdez-Valdivia, J.A. García: Mapping citation patterns of book chapters in the Book Citation Index, J. Informetr. **7**(2), 412–424 (2013)

27.108 SENSE-Research School for Socio-Economic and Natural Sciences of the Environment: SENSE ranking of academic publishers, http://www.sense.nl/qualityassessment (2009)

27.109 A. Zuccala, M. van Someren, M. van Bellen: A machine-learning approach to coding book reviews as quality indicators: Towards a theory of mega-

Part D | 27

citation, J. Assoc. Inf. Sci. Technol. **65**(11), 2248–2260 (2014)

27.110 A. Zuccala, F. Verleysen, R. Cornacchia, T. Engels: Altmetrics for the humanities: Comparing Goodreads reader ratings with citations to history books, ASLIB Proceedings **67**(3), 320–336 (2015)

27.111 J. Gorraiz, C. Gumpenberger, P.J. Purnell: The power of book reviews: A simple and transparent enhancement approach for book citation indexes, Scientometrics **98**(2), 841–852 (2014)

27.112 A. Zuccala, R. Guns: Comparing book citations in humanities journals to library holdings: Scholarly use versus "perceived cultural benefit" (RIP). In: *Proc. 13th Int. Soc. Sci. (ISSI) Conf.* (AIT Austrian Institute of Technology GmbH Vienna, Vienna 2013) pp. 353–360

27.113 D. Torres-Salinas, N. Robinson-Garcia, J. Gorraiz: Filling the citation gap: Measuring the multidimentsional impact of the academic book at institutional level with PlumX, Scientometrics **113**(3), 1371–1384 (2017)

27.114 B. Flyvberg: *Making Social Science Matter: Why Social Inquiry Fails and how it can Succeed Again* (Cambridge Univ. Press, Cambridge 2001)

Alesia Zuccala

Department of Information Studies
University of Copenhagen
Copenhagen, Denmark
a.zuccala@hum.ku.dk

Alesia Zuccala is an Associate Professor at the Department of Information Studies, University of Copenhagen. She graduated with a PhD from the Faculty of Information Science at the University of Toronto (Canada) in 2004. Her research specialization is in the field of bibliometrics, scholarly communication, and research evaluation, and in recent years she has focused primarily on the evaluation of research outputs across the humanities.

Nicolas Robinson-Garcia

School of Public Policy
Georgia Institute of Technology
Atlanta, GA, USA
elrobinster@gmail.com

Nicolas Robinson-Garcia is a postdoctoral researcher at the School of Public Policy from the Georgia Institute of Technology in the United States. He holds a PhD in Social Sciences from the University of Granada (Spain). He is currently interested in the development of quantitative methods to analyze scientific mobility flows as well as the analysis of new data sources and methods to trace societal engagement of researchers.

28. Scholarly Twitter Metrics

Stefanie Haustein

Twitter has unarguably been the most popular among the data sources that form the basis of so-called altmetrics. Tweets to scholarly documents have been heralded as both early indicators of citations and measures of societal impact. This chapter provides an overview of Twitter activity as the basis for scholarly metrics from a critical point of view and equally describes the potential and limitations of scholarly Twitter metrics. By reviewing the literature on Twitter in scholarly communication and analyzing 24 million tweets linking to scholarly documents, it aims to provide a basic understanding of what tweets can and cannot measure in the context of research evaluation. Going beyond the limited explanatory power of low correlations between tweets and citations, this chapter considers what types of scholarly documents are popular on Twitter, and how, when and by whom they are diffused in order to understand what tweets to scholarly documents measure. Although the chapter is not able to solve the problems associated with the creation of meaningful metrics from social media, it highlights particular issues and aims to provide the basis for advanced scholarly Twitter metrics.

28.1 Tweets as Measures of Impact

Enabled by the digital revolution, the open access and open science movement, big data and the success of social media have shaken up the scholarly metrics landscape. Academic careers are no longer shaped only by peer-reviewed papers, citation impact and impact factors; university managers and funders now also want to know how researchers perform on social media and how much their work has impacted society at large.

Bibliometricians have started to adapt to the policy pull and technology push and expanded their repertoire of scholarly metrics to capture output and impact beyond the ivory tower, so far that some speak about a scientometric revolution [28.1]. Metrics are no longer restricted to formal parts of communication but expand beyond the borders of the scholarly community [28.2]. Similarly to how the Science Citation Index formed the field of bibliometric research and research evaluation, the altmetrics, or more precisely the social-media metrics landscape is being heavily shaped—if not entirely driven—by the availability of data, in particular via automated programming interfaces (APIs) [28.3, 4].

Twitter has arguably been at the epicenter of the earthquake that has shaken up the scholarly metrics landscape. The majority of altmetrics research has either focused on or included Twitter (see [28.5] for a review of the literature). Following a general definition of scholarly metrics which include activity on social media [28.3], scholarly Twitter metrics are defined as "indicators based on recorded events of acts [on Twitter] related to scholarly documents [. . .] or scholarly agents [. . .]" [28.3, p. 416].

Although findings of an early study had suggested that tweets were a good early indicator of citations for papers published in the *Journal of Medical Inter-*

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_28

net Research [28.6], the generalizability of this claim was refuted by low correlations reported by more representative studies [28.7, 8]. Low correlations between tweets and citations did, however, spark hopes that Twitter activity was able to reflect impact on users and use beyond citing authors—a new type of previously unmeasurable impact—possibly on society at large. Twitter's popularity in the altmetrics realm has essentially been caused by two factors, which are both heavily influenced by technology and the data push and policy pull described above:

1. A significant number of scholarly articles are shared on Twitter, producing a measurable signal.
2. Twitter is a social media platform created for and used by a wide and general user base, which theoretically has the potential to measure impact on society at large.

As tweets represent an "unprecedented opportunity to study human communication and social networks" [28.9, p. 1814], Twitter is being used to analyze a variety of social phenomena. Centering on either the message (i. e., the tweet, its content and associated metadata) or social connections (i. e., the network of follower-followee relations), tweets have been used to show discussions during upcoming elections, how people communicate during natural disasters, political upheaval, cultural events and conferences and have even been used to predict election outcomes and the stock market [28.10, 11].

While only a small share of academics use Twitter for scholarly communication [28.12, 13] and to diffuse scientific publications [28.14], more than one fifth of recent journal articles are being tweeted [28.15], which adumbrates that it is non-academics who engage with scholarly publications on Twitter. At this point, social media-based indicators have flourished rather as vanity measures—culminating in a tongue-in-cheek metric called the Kardashian Index [28.16]—than as validated indicators of societal impact. Even though altmetrics have left their mark on the scholarly publishing and metrics landscape [28.17], they have not (yet) established themselves within the reward system of science, where citations remain the only hard currency [28.18, p. 1523]:

> Neither Twitter mentions nor Facebook 'likes' are, for now at any rate, accepted currencies in the academic marketplace; you are not going to get promoted for having been liked a lot, though it may well boost your ego.

Still, almost all big publishing houses now report some form of article-level metric based on social media activity, including tweets. Despite the lack of validation and a clear definition regarding the type of impact measured, the number of tweets are thus already used as scholarly metrics "in the wild" [28.19].

This chapter aims to contribute to the understanding of Twitter and Twitter-based metrics with a particular focus on their potential and limitations when applied as scholarly metrics. To provide some context for the meaning of scholarly impact measures derived from tweets and Twitter activity, this chapter describes Twitter's role in scholarly communication. It depicts how Twitter is used in academia and how scholarly contents are diffused and discussed in tweets. The chapter provides an overview of the literature of Twitter use by the scholarly community and scholarly output on Twitter. The latter part is supported by empirical results based on an analysis of 24 million tweets mentioning scientific papers captured by the data provider Altmetric.com. Both the review of the relevant literature and the patterns extracted from the Twitter data are intended to contribute to the understanding of what type of scholarly contents are diffused on Twitter, who is diffusing them, when and how. This will help to assess Twitter metrics as valid impact indicators and to interpret their meaning.

28.2 Twitter in Scholarly Communication

Twitter launched in 2006 as a public instant messaging service and evolved from an urban lifestyle social network, where users would update their friends about what they were doing, to a platform for communicating news and events used by 500 million users worldwide, or 23% of US adults online [28.20]. Although other microblogging platforms (e. g., Sina Weibo, Tumblr, Plurk) exist, tweeting has become a synonym (and preferred term) for microblogging and Twitter the most popular service. Until recently, Twitter constrained microposts to a maximum length of 140 characters, a restriction that originates from the 160-character limit of text messages. In 2017, the restriction was increased to 280 characters. Users can follow each other and create user lists to manage the updates they receive from other Twitter users. Similarly to regular blogs, microblogs are ordered sequentially in reverse chronological order and, due to their brevity, usually appear more frequently [28.21], making Twitter the "most dynamic and concise form of information exchange on social media" [28.22, p. 5]. While the brevity of tweets is seen as a restriction by some, others perceive it as a particular advantage [28.23, p. 10]:

> The brevity of messages allows [tweets] to be produced, consumed, and shared without a significant amount of effort, allowing a fast-paced conversational environment to emerge.

Tweets have three major specific affordances, which facilitate communication on the platform: retweets (RTs), user mentions (@mentions) and hashtags (keywords following #). All of these functions originated within the Twitter user base and were eventually adopted by Twitter, representing a co-creation of functions by users and developers. Twitter provides three main levels of communication: interpersonal communication on the micro level, meso-level exchanges of people who are directly connected through their network of followers and followees, and hashtag-centered macro-level communication which enables exchanges among all Twitter users with common interests [28.24].

In academia, Twitter is used to disseminate and discuss scholarly outputs and other relevant information; maintain collaborations or find new ones; as a virtual "water cooler" [28.25, p. 347] for social networking with colleagues; to increase student participation in teaching; as a back channel at scientific conferences to foster discussions among conference attendees and those who participate remotely; and to increase visibility and reach wider audiences [28.13, 25–29].

28.2.1 Twitter Uptake

Twitter is used by various stakeholders in the scholarly community, including individual researchers and academics, journals and publishers, universities and other academic institutions, as well as at scholarly conferences. From the perspective of using Twitter activity as the basis for scholarly metrics, it is essential to know Twitter uptake in academia, as it informs about biases and differences between disciplines and other user demographics, which may have a direct effect on derived metrics.

Scholars on Twitter

In the scholarly context, Twitter use by academics lags behind its uptake among the general public. Although the majority of researchers are aware of the platform, most do not make use of it in a professional context, giving it the reputation of a hype medium in academia [28.13, 30–32]. A certain reluctance in academia to use Twitter might be caused by its perception as a shallow medium that is used to communicate "pointless babble" [28.33, p. 5] rather than informative content [28.10]. Described as "phatic" [28.34, p. 396], Twitter is less about what people tweet rather than how they are connected.

Reported Twitter uptake shows extreme variations depending on user demographics—in particular disciplinary orientation—and when a study was conducted; it usually stays behind use of other social media. For example, a survey among 2414 researchers conducted in 2010 demonstrated that while more than three quarters used social media, less than one fifth were on Twitter [28.12]. A more recent study conducted by *Nature* also showed that Twitter was among the least used social media platforms in academia: while almost half of 3027 science and engineering researchers regularly used ResearchGate, only 13% regularly visited Twitter [28.13]. At almost one quarter of regular users, Twitter uptake was higher among the 482 social science and humanities scholars participating in the same survey [28.13].

Depending on the sample and when the survey was conducted, Twitter uptake varied heavily between a few percent to more than one third of surveyed scholars using Twitter, which calls the representativeness of findings into question. Moreover, surveys vary in terms of whether or not they differentiate between general Twitter uptake, Twitter use for scholarly communication and professional purposes or active versus passive use, which further complicates comparison and generalization of findings regarding Twitter uptake in academia. Twitter represented the social media tool with the highest difference between awareness and use. Although known by 97% of university staff in Germany, as few as 15% used Twitter and 10% used Twitter in a professional context [28.32]. A similar use-to-awareness ratio was found by other studies, for example, at Finnish universities [28.31] or researchers surveyed by *Nature* [28.13], and academic staff in Germany [28.35].

Most studies found self-reported Twitter use in academia at around 15%; an uptake of 13–16% was reported for surveys based on 215 health services and policy researchers [28.36], 454 geographers [28.37], 1058 UK academic staff [28.38] and 3027 scientists and engineers [28.13], while uptake was lower (7–10%) for academics in Germany [28.35] and the UK [28.12, 39]. Although 18% used it, Twitter was the least popular social media tool of 345 European scholars [28.40]. The highest Twitter use was reported for a survey of 126 Finnish university staff at 23% [28.31], 1910 professors at US universities at 32–23% for professional purposes [28.41], 382 urologists attending a conference at 36% [28.42] and 71 participants of the 2010 *Science & Technology Indicators* (STI) conference in Leiden at 44% [28.43].

As an alternative to determining Twitter uptake through self-reported use in surveys, studies have also assessed the extent of scholarly microblogging based on Twitter activity of scholars. Identifying scholars on Twitter is challenging, as the 160-character Twitter bio and the provided user name are often the only basis for identification. Most studies thus search for Twitter users based on a list of names of academics [28.44–48] or apply snowball sampling starting from a set of known scholars on Twitter [28.49–51]. Searching for a list of 8038 US and UK university staff, *Priem* et al. [28.47] found Twitter accounts for 2.5% of them. Although the authors admit that their study underestimated Twitter use, it reflected that the microblogging platform is not popular in academia and thus confirms the findings by most surveys. Investigating Twitter use in the scientometric community, *Bar-Ilan* et al. [28.52] found Twitter profiles for 9 of the 57 presenters at the 2010 STI conference.

Other than searching for known scholars on Twitter, some studies try to extract information from Twitter to identify scholarly users. The most common approach is to classify users based on searching for specific words in the Twitter bio. Retrieving users whose Twitter bio contained words such as *university*, *PhD* or *professor*, *Barthel* et al. [28.53] identified scientists' Twitter accounts with a precision of 88%. False positives contained university accounts or those of non-academic staff at research institutions. Recall cannot be determined in such studies as the number of false negatives, that is the scientists on Twitter who do not include any of the queried keywords in their self-descriptions, remains unknown.

Altmetric also applies a keyword-based approach to categorize Twitter users as *scientists*, *science communicators (journalists, bloggers, editors)*, *practitioners (doctors, other healthcare professionals)* and *members of the general public*. It should be noted that Altmetric's *general public* category includes all users that cannot be classified as belonging to any of the other three groups and is therefore not a good indicator of how much an article has been tweeted by members of the general public. An obvious limitation of the keyword approach is that it is unable to capture scholars who do not identify themselves as such or who do not use the terminology or language covered by the list of keywords. However, many scholars seem to reveal their professional personas on Twitter. Ninety percent of doctoral students funded by the Canadian Social Sciences and Humanities Research Council (SSHRC) identified as academics on Twitter [28.48], 87% of surveyed US university professors claimed to mention both their professional title and place of work in their Twitter profiles [28.41] and 78% of Twitter users who self-identified as a physicians used their full names [28.49]. This willingness to reveal their scholarly identities on Twitter suggests that scholars make use of the microblogging platform in a professional context at least to some extent.

Ke et al. [28.54] took advantage of crowdsourced Twitter lists to identify scholars. Based on a method introduced by *Sharma* et al. [28.55], they identified scientists on Twitter with an approach based on membership in scientific Twitter user lists. Other studies have estimated Twitter activity by scholars by analyzing users who engage with scholarly content on Twitter. *Hadgu* and *Jäschke* [28.56] applied machine learning to automatically identify scholars on Twitter based on a training set of users whose tweets contained a computer science conference hashtag, while others selected users who have tweeted scientific papers [28.51, 57–60]. Since the latter type of studies focuses on categorizing who is tweeting about scholarly contents rather than estimating Twitter uptake in academia, these studies are discussed in more detail below.

Scientific Conferences

Twitter does particularly well in fostering communication among people participating in shared experiences [28.10], which may be why tweeting at scholarly conferences has been one of the earliest and most popular uses of Twitter in academia. Almost every scientific conference today has a specific hashtag to connect attendees and those interested but not able to attend in person, thus expanding the conference audience to include remote participants [28.61–63]. Apart from increasing the visibility of presentations, tweeting at scientific conferences has introduced another level of communication, creating back-channel discussions online among participants complementing presentations and discussions which take place at the meeting. Conference tweets usually directly refer to presentations and discussions during sessions and sometimes summarize key take-away points [28.64–66]. Other motivations for tweeting at a scientific conferences were to share information and learn about discussions in parallel sessions, networking with others and feeling a sense of connectedness, as well as note-taking [28.65]. A significant number of tweets associated with two medical conference were uninformative or promotional [28.45, 67].

Due to the ease of collecting tweets with a particular hashtag, as well as Twitter's relative popularity in the context of scientific conferences, there are countless studies analyzing scholarly Twitter use based on tweets with conference hashtags [28.62–72].

Similar to the overall uptake among scholars, Twitter activity at conferences differs among disciplines as well as individual conferences, and has increased over the years. Overall, only a small share of conference participants contributed to discussions on Twitter: Less than 2% of attendees of the *American Society of Nephrology*'s 2011 conference [28.45] and less than 3% of participants of the 2012 *Winter Scientific Meeting of the Association of Anesthetists of Great Britain and Ireland* tweeted [28.65]. Another medical conference in 2013 reported higher Twitter engagement, as 13% of conference attendees tweeted using the conference hashtag [28.66]. Longitudinal studies also observed an increase in Twitter activity at conferences over the years [28.64, 66, 69]. For example, 2% of conference participants tweeted at the *2010 Annual Meeting of the American Society of Clinical Oncology*, while 5% contributed to conference tweets in 2011. Similarly, the number of tweets nearly doubled from 4456 tweets in the first to 8188 in the following year [28.64]. A similar increase was observed for the 2011 and 2012 annual meetings of the *Radiological Society of North America* [28.69]. Conference-related discussions on Twitter are not restricted to in-person attendees. In fact, at some conferences the majority of Twitter users only participate remotely [28.62].

Just as with other social media and information in general, tweeting activity is usually heavily skewed with a few users contributing the majority of tweets at conferences [28.64, 66, 67]. Tweeting about a conference has been shown to lead to an increase in the number of followers regardless of attending in-person or remotely. Follower counts grew particularly for speakers and in-person attendees, while the number of followers grew least for remote participants [28.62]. Most organizers of scientific conferences embrace the potential of increasing visibility and outreach and thus encourage tweeting through a conference-specific hashtag. Some also specifically display conference-related tweets in real time and thus make tweeting activity visual to participants who are not on Twitter [28.62, 63, 68, 70].

Journals and Publishers

Twitter's technological features afford direct connections and two-way conversations between users, changing what was traditionally known as a unidirectional sender-audience relationship. Opposed to traditional publishing and mass media, Twitter has given rise to personal publics of audiences [28.73]. This direct link between the sender and receiver has changed the relationship with audiences; for example, musicians use Twitter to market their own brand and respond to @replies from fans to seek out in-person interactions [28.74]. TV audiences turn Twitter into a virtual lounge room when they connect with other users discussing TV events in real time [28.75]. Similarly, discussions of scientific publications can now happen publicly, when readers share their opinions on Twitter. A specific use case are Twitter journal clubs, an adaption of small-group in-person journal clubs that are particularly common in the medical sciences [28.76–80]. Twitter journal clubs are used to discuss and review recent publications and educate researchers and practitioners; in the medical sciences they also have the advantage over their offline predecessors to directly involve patients [28.77]. Often these journal clubs are initiated or at least supported by journals to promote their publications. A journal club initiated by a gynecology journal showed that discussing papers and making them freely available has boosted their Altmetric scores [28.76]. Twitter journal clubs also motivated authors of discussed papers to create Twitter accounts [28.78].

With journals and authors on Twitter, readers can get in touch directly and involve them in discussions using @mentions, tearing cracks in the wall of traditional gatekeeping, as "Twitter makes it possible to directly connect journal readers at various stages of training with authors and editors" [28.77, p. 1317]. Many journals and publishers have started to use Twitter as a marketing instrument to increase online visibility and promote published contents. These accounts can be used to create a personalized audience relationship and to foster interaction among readers. Similar to the mix of professional and personal interactions by academics on Twitter, the lines between scholarly communication and marketing campaign are blurred for accounts maintained by journals and publishers. Almost half of the 25 general medicine journals with the highest impact factor in 2010 had a Twitter presence [28.81], while Twitter uptake was lower for other sets of journals: 24% of 33 urology journals [28.82], 2 of the top 10 ophthalmology journals [28.83], 16% of 100 *Web of Science* (WoS) journals [28.84] and 14% of 102 journals specialized in dermatology [28.85] maintained an account. As most of these studies focused on the top journals according to the journal impact factor, Twitter uptake might be biased towards high-impact journals and slightly lower when including others. The variation suggests similar differences between disciplines as observed for Twitter uptake by individual scholars and conferences.

While most journal accounts are used to share articles and news [28.86] and often tweet the article title [28.87, 88], some journals have incorporated tweeting into the formal communication process. In addition to regular abstracts, they ask authors to write so-called tweetable abstracts that meet the 140-character restrictions, which are used to attract readers on Twitter [28.44]. Twitter even interfered with the journal's role in scholarly communication, when a genomics paper was criticized and corrected results posted in a tweet, leading to a conflict with the authors of the criticized paper [28.89].

Even if a journal is not represented by a proper Twitter account, it is likely that its publisher is. *Zedda* and *Barbaro* [28.86] found that Twitter adoption was particularly common among 76 publishers in science, technology and medicine; 89% had official Twitter accounts, exceeding the presence on any other social media platform, and 74% had embedded tweet buttons that allowed readers to directly share publications on Twitter. Promotion of publications by publishers seem to be welcomed by authors, as a survey by *Nature Publishing Group* revealed that almost one fifth of authors would consider it a very valuable service if publishers promoted papers using marketing and social media [28.90].

As shown in the analysis of Twitter accounts diffusing scientific articles below, accounts maintained by journals and publishing houses are responsible for a significant amount of tweets mentioning scientific articles. Once Twitter metrics are being used to evaluate journal impact, these types of self-tweets might be considered as a type of *gaming* in a manner similar to journal self-citations and citation cartels to boost the impact factor [28.91, 92]. With publishers invested in the success of their journals, *tweet cartels* and *tweet stacking* in analogy to their citation equivalents are easily conceivable and even easier to implement. While the WoS excludes journals from the *Journal Citation Reports*, which have been caught increasing their impact factors artificially, companies like Altmetric.com and Plum Analytics do not (yet) intervene in such self-promotional activity.

Universities and Academic Libraries

Scholarly institutions are affected by Twitter's impact on academia on two levels: they exploit the microblogging platform to increase their visibility (and that of their members) and provide guidelines and recommendations for their members to navigate the new communication space. The Association of American University Professors updated their report on *Academic Freedom and Electronic Communication* [28.93] in reaction to a university rescinding a tenure-track job offer to an English scholar who had made an anti-Semitic comment on Twitter [28.94]. The updated report emphasized that professors enjoy academic freedom even when they comment on social media and particularly addressed the blurring of boundaries between private and professional opinions on social media. It stressed how, in this new context, comments are particularly prone to be misunderstood and misinterpreted, as they are often taken out of context [28.93, p. 42]:

> Electronic communications can be altered, or presented selectively, such that they are decontextualized and take on implicit meanings different from their author's original intent. With the advent of social media, such concerns about the widespread circulation and compromised integrity of communications that in print might have been essentially private have only multiplied further.

The report further recommends that universities and other academic institutions, along with their staff, develop policies that address the use of social media. In general, academic institutions lack specific social media guidelines or address social media in policies. Although more and more institutions adopt specific policies [28.95], only half of US doctorate-granting universities had a social media policy, while rates were even lower for other universities and colleges. At the same time, Twitter was specifically mentioned in more than 80% of policies [28.96].

The majority of university Twitter accounts apply a so-called megaphone model of communication, where news and information concerning the institution are broadcast following a traditional communication model [28.97, 98]. Universities use Twitter for public relations, dissemination of news and events, and as recruitment [28.97]. Ninety-four percent of 474 US university admission officers reported that their institution had a Twitter account [28.99] and 96% of the websites of 100 US colleges linked to Twitter [28.100]. On the departmental level, it was less common to be represented with an organizational account, as only 8% of 183 US radiology departments had a Twitter presence [28.101]. Twitter was also commonly used for faculty development at medical schools [28.102]. Analyzing the Twitter activity of 29 Israeli universities and colleges, *Forkosh-Baruch* and *Hershkovitz* [28.103] found significant differences between both types of institutions. Colleges were more likely than universities to post social tweets. While almost half of the tweets by universities focused on research conducted elsewhere, colleges focused more on reporting the work by its own researchers.

Twitter was also frequently used by academic libraries as a marketing instrument, to communicate with patrons, to announce new resources and promote services [28.104–106]. On par with Facebook at a 63% adoption rate, Twitter was the most commonly used social media platform among 38 surveyed academic libraries from different countries [28.107], while all of the 100 US university libraries analyzed by *Boateng* and *Quan Liu* [28.104] maintained a Twitter presences. In Canadian academic libraries, Twitter adoption was lower at 47% [28.108].

28.2.2 Twitter Use

Apart from Twitter uptake, scholarly Twitter metrics are further influenced by *how* Twitter is used. Academics use social media to share information, for impression management and to increase their visibility online, to network and establish a presence across platforms, to request and offer help, expand learning opportunities, or simply to be social [28.25]. Twitter specifically was used mostly to tweet work-related content, discover peers working on similar research, follow research-related discussions and get recommendations for papers [28.13].

One central motivation for tweeting among scholars is to communicate and explain their work to laypeople. As many science communicators are active on Twitter, they help to bridge the gap between the scholarly community and the general public. Science communicators were the largest user group of 518 Twitter users mentioned in tweets by 32 astrophysicists [28.50]. An evolutionary biology professor valued Twitter to communicate his work to the general public [28.61, p. 453]:

> Twitter and regular blogging are more effective than anything else I do to publicize a paper, which was really surprising to me [...]. If you do it right, Twitter is an effective way of telling people about your work.

However, most researchers still preferred traditional media over Twitter to promote their research [28.109].

Even when identifying professionally on Twitter, a large share of tweets by scholars are not related to their work or academia in general [28.13, 32, 41, 110, 111]. The *Nature* survey found that 21% of scientists who used Twitter regularly did not use it professionally and 28% said that they never posted content about their work [28.13]. *Bowman* [28.41, 112] reported that, while 29% of American university professors used Twitter strictly in a professional way and 42% used it for both for personal and professional reasons, the vast majority of tweets were coded as personal (78%) rather than professional (19%). Again, large variations can be observed between disciplines as well as individual Twitter users [28.42, 47–49, 51, 111]. Examining more than half a million tweets from 447 researchers, *Holmberg* and *Thelwall* [28.51] found that less than 4% of tweets were classified as scholarly communication and results varied between disciplines ranging from less than 1% for sociology up to 34% for biochemistry. A study on emergency physicians' tweeting behavior found that 49% of their tweets were related to health or medical issues, 21% were personal, 12% self-promotional and 3% considered unprofessional as they contained profanity, were discriminatory or violated patient privacy [28.49]. In a sample of tweets by funded doctoral students in the social sciences and the humanities in Canada, 4% of tweets were related to their thesis, 21% to the discipline and 5% to academia in general, while 70% of tweets were coded as non-academic [28.48]. Personal use also prevailed among 382 urologists [28.42].

These findings highlight that even when scholars identify professionally on Twitter and use the platform for scholarly purposes, many tweets will be irrelevant to scholarly communication and should thus be excluded from a scholarly indicator perspective [28.21, p. 98]:

> The lack of a dividing line between scientists and non-scientists, as well as the great variety of topics that even scientists tweet about mean that Twitter is not comparable to the orderly world of science publishing, where every piece of information is assumed to be relevant. Instead, a typical user's timeline is likely to be populated both by scholarly content and personal remarks, more or less side by side.

Due to their brevity and the fact that when analyzing tweets they are often taken out of context, categorizing tweet content is as difficult as it is to classify Twitter users [28.41]. Distinguishing between scientific and non-scientific tweets is especially challenging [28.51].

Large variations can also be found between individual tweeters in terms of how often they tweet. A group of astrophysicists analyzed by *Haustein* et al. [28.110] tweeted, on average, between 0 and 58 times per day. Tweets to scientific papers have been shown to peak shortly after their publication and decay rapidly within just a few days. For example, 80% of *arXiv* submissions received the largest number of tweets the day after they were published [28.113]. Similarly, the tweeted half-life was 0 days for papers published in the *Journal of Medical Internet Research* [28.6], and 39% of a sample of tweets linking to a scholarly document referred to those published within one week before [28.14]. Determining the delay between publication and first tweet as well as half-lives on Twitter is, however, challenging due to the ambiguity of publication dates [28.114].

Since not all Twitter use culminates in a tweet, a large share of activity remains invisible and thus unmeasurable. In fact, passive use prevailed among UK doctoral students on Twitter [28.30] and seems to be common for scholarly use of social media in general. While most academics access and view information, only a minority actively contributes by creating content on social media [28.38, 39]; less than 2% of 1078 UK researchers surveyed actively contributed daily [28.38].

Tweeting Links

The most frequent use of Twitter among researchers in higher education was to share information, resources or media [28.25]. A survey among US university professors revealed that embedding URLs was the most commonly used Twitter affordance. Half of the survey participants claimed to tweet links either sometimes, mostly or always [28.41]. Links are a common way to send more information than 140 characters would fit. Addressing the length limitation, a scholar explained [28.61, p. 453]:

> It is a double-edged sword. The majority of my tweets are pointers to other resources, so there is a headline—an enticement in other words—and a link to the resource. You don't need more than 140 characters for that.

Weller and *Puschmann* [28.115] refer to links in tweets as "external citations" [28.115, p. 2]. Studies about scholars on Twitter show that they make frequent use of tweeting URLs, as the share of tweets with links exceeds that observed for general Twitter users [28.23, 75]. About one third of 68 232 tweets sent by 37 astrophysicists [28.110] and 38% of 22 258 tweets posted by Canadian social sciences and humanities doctoral students contained links [28.48]. Tweeting links was even more popular among scholars studied by *Weller* and *Puschmann* [28.115] and emergency physicians analyzed by *Chretien* et al. [28.49], as 55% and 58% of their tweets, respectively, contained URLs. Links were much more common when a sample of 445 US professors tweeted professionally: 69% of professional tweets contained URLs, while only 15% of personal tweets did [28.41]. Tweets with the #www2010 and #mla09 conference hashtags linked to a website in 40% and 27% of the cases, repeating each unique URL less than three times [28.63].

Priem and *Costello* [28.14] found that 6% of a sample of 2322 tweets by academics containing a URL mentioned a scholarly publication, 52% of which were first-order and 48% second-order links (i. e., via another website) to the document. Similarly, *Holmberg* and *Thelwall* [28.51] found that scholarly tweets frequently contained a link to scholarly publications via a blog post about the paper. First-order links were significantly more likely to refer to open access articles [28.14]. Tweets containing the #iswc2009 conference hashtag linked to applications (e. g., online services or research projects; 31%), the conference website (21%), blog posts (12%), slideshows (12%) and publications (9%) [28.71]. Blogs were the most common linked resources in conference tweets analyzed by *Weller* and colleagues [28.63], while news websites were a frequent link destination of tweets sent by Canadian doctoral students, even when discussing scholarly topics [28.48].

When linking to scholarly papers, tweets often contained the paper title and rarely expressed any recommendation or sentiment [28.87, 88]. The great majority of tweeted articles were published very recently [28.6, 14, 51]. According to surveys asking about motivations for using social media, finding relevant publications and staying up to date with the literature was found to be a frequent, albeit passive, use of Twitter [28.13]. A Columbia university professor in biology and chemistry describes how they used Twitter to be alerted about the literature [28.61, p. 452]:

> Sometimes four or five people I follow will mention a paper that I did not come across and I will look it up. I think I am much more up to date on science literature since I started following Twitter.

A study by *Tenopir* et al. [28.38] found that academics on Twitter read more scholarly publications, which seems to confirm the use of Twitter as a publication alert service. At the same time, as the analyses by *Priem* and *Costello* [28.14] and *Letierce* et al. [28.71] show, the share of tweets linking to academic papers is low, suggesting a rather passive use: scholars follow links to tweeted articles but do only infrequently distribute them themselves. This suggests that a significant part of Twitter use cannot be captured by scholarly Twitter metrics.

Retweets

Retweets represent a specific form of diffusing information, as users forward messages sent by others. As such, they do not represent an original contribution by the retweeting user. Since retweets directly quote another users text, they can be seen as "internal citations" [28.115, p. 3] on Twitter. An analysis of retweets demonstrates how information circulates within a specific user community [28.116]. A common disclaimer that *retweets do not equal endorsements*, adapted from early Twitter use by journalists, emphasizes that tweets are forwarded to increase information diffusion. Once a frequent part of Twitter bios, the disclaimer has now been established as common sense and is no longer needed [28.117].

As sharing information is one of the main motivations for scholarly Twitter use, retweeting is likely to be common among tweeting academics. Conference participants interviewed by *Letierce* and colleagues retweeted "tweets that are close to their interest or tweets that speak about their own work or research project" [28.71, p. 7]. Studies showed that retweeting is less common than other affordances used by scholars on Twitter, but exceeds expectations of a random sample of tweets in 2009, in which as few as 3% were retweets [28.23]. Between 15% and 20% of tweets sent at scientific conferences were retweets [28.71]; similarly, 15% of 68 232 tweets by a group of 37 astrophysicists were retweets [28.110]. Retweeting was more common among 28 academics analyzed by *Priem* and *Costello* [28.14], as 40% of tweets were retweets. Similarly, 37% of 43 176 tweets by Canadian doctoral students were retweets, while 10% of their tweets were retweeted [28.48]. At the same time, the share of retweets at a radiology conference was 60% [28.69]. Asked about Twitter affordance use, 14% of US professors said that they mostly or always retweeted, and 34% that they sometimes retweeted [28.41]. These professors were more than twice as likely to retweet when their tweets were classified as professional rather than personal [28.41]. These differences again demonstrate that tweeting behavior differs depending on who is tweeting and in what context.

The majority of retweets sent by scholars from ten disciplines contained links, while conversational tweets (i. e., @mentions) were less likely to contain links [28.51, p. 1035]:

> This clearly shows that researchers [...] frequently share web content and forward information and content they have received from people they follow on Twitter, while links are not that often shared in conversations.

Links to papers were significantly less likely to be retweeted: while 19% of tweets with links to scholarly publications were retweeted, the retweet rate was twice as high in the overall sample of tweets analyzed by Priem and Costello [28.14]. This is in contrast to a random sample of more than 200 000 tweets, over half of which contained a URL [28.23]. A random sample of 270 tweets linking to scientific journal articles found that many were modified retweets of tweets originating from the journal's own Twitter account [28.88].

From the perspective of scholarly metrics, a distinction should arguably be made between tweets and retweets, as the latter reflects a rather passive act of information sharing [28.118]. Although with each retweet the visibility of the tweet and the information it contains (e. g., the link to a publication) increases, retweets represent diffusion of information rather than impact. As retweeting requires even less effort—as little as one click since the implementation of the retweet button—than composing an original tweet, Twitter metrics should distinguish between tweets and retweets to reflect these different levels of user engagement [28.119].

Followers, @mentions and @replies

Scholars on Twitter actively seek new connections and connect others [28.25]. Twitter is built in a way that information is spread via user networks. This means that users receive updates from those they follow and diffuse their messages to those they are followed by, creating a personal public [28.73]. Selecting who to follow and who one is followed by are thus essential to communicating on Twitter. Twitter users build a reputation based on both their number of followers and followees. A parallel can be drawn to authors' citation identity and citation image based on which authors one cites and is cited by [28.120, 121].

The purpose of expanding one's social network and finding peers was apparent in a study of 632 emergency physicians on Twitter: those who included work-related information in their Twitter bios had more followers and the most influential users in the network were connected to at least 50 other emergency physicians [28.122]. Similarly, conference participants on Twitter saw a significant increase in their number of followers, particularly when they were also presenting [28.62]. The average number of followers of 260 physicians analyzed by *Chretien* et al. [28.49] was 17 217 with a median of 1426, indicating the typically skewed distribution of followers among Twitter users.

In addition to broadcasting one's message via retweets on a meso level of communication, Twitter users can also directly address users with @mentions or @replies on an interpersonal level. Both replies and mentions thus represent a particular type of tweet that focuses on conversation rather than broadcasting [28.123]. Like most Twitter affordances, these types of tweets were also developed by the user base before being implemented by Twitter. While @replies happen in response to a tweet and are only visible on the timeline of the tweeter who sent the original tweet, @mentions refer to tweets that contains another user's Twitter handle, which triggers a notification to inform them about being mentioned [28.24]. About one third of tweets in 2009 included another user name [28.23]. A random sample of tweets without an @mention were mostly about the tweeting user's experience, while those with an @mention were more likely about the addressee. In fact, more than 90% of @mentions functioned to address another user, while 5% worked as a reference [28.123].

The great majority of tweets sent by social sciences and humanities doctoral students referred to other Twitter users, as 72% of the 43 176 tweets sent contained other user names [28.48]. Conversational tweets were also popular among a group of tweeting astrophysicists. Of the 68 232 tweets, 46% were @replies or @mentions (61% including RTs), making it the most frequently used Twitter affordance. Conversational tweets were particularly common among those who tweeted regularly or frequently [28.110]. Most of these mentions referred to science communicators (24%), other astrophysicists (22%) or organizations (13%) on Twitter [28.50].

Conversational tweets hardly contained links [28.51], and among tweets linking to publications, only 8% were @replies [28.14]. Opposed to hashtag use, retweeting and embedding URLs, @mentions were the only affordance that were less likely to occur in professors' professional tweets (56%) than those identified as personal (67%), which suggests that when professors discuss their work, they are less likely to address or reference other users directly than when they tweet about private matters. However, mentioning other users was still more common in their professional tweets than retweeting and using hashtags [28.41].

Hashtags

Similar to retweets and @mentions, hashtags are a user-driven Twitter affordance. Hashtags are keywords following the #-sign, which facilitate connections between users interested in the same topics. Conversations revolving around hashtags represent the macro layer of Twitter communication [28.24]. Holmberg and colleagues suggest that "hashtags may resemble the traditional function of metadata by enhancing the description and retrievability of documents" [28.50, p. 3].

Hashtag use seems to be less common than that of other Twitter affordances among academics. Sixty-one percent of surveyed US-American university professors declared that they rarely or never used a hashtag [28.41]. This might be because scholars are either less familiar with this Twitter-specific affordance or they do not wish to expand conversations beyond their personal publics defined by their follower networks. Although actual hashtag use was low by the US professors analyzed by *Bowman* [28.41], they were more likely to use hashtags in their tweets identified as professional (28%) than those coded as personal (17%).

An early large-scale study found that among a random sample of 720 000 tweets, only 5% contained a hashtag [28.23]. Almost one quarter of tweets by astrophysicists contained a hashtag [28.110], but hashtag use varied significantly among different clusters of the follower network [28.50]. The same share of hashtags (25%) was found for tweets of SSHRC-funded doctoral students; on average each hashtag was mentioned 2.8 times [28.48]. It is problematic to infer hashtag use from most other studies on scholarly tweets, as data collection itself is often based on a specific hashtag, such as a conference hashtag [28.63, 71].

28.2.3 Reluctance Against and Negative Consequences of Using Twitter

When using tweets as the basis to measure scholarly impact of any sorts, it is essential to consider who is not on Twitter and why academics might be reluctant to join the microblogging platform. Twitter is a platform where "content is not king" [28.34, p. 395] and has been perceived as "shallow media, in the sense that it favors the present, popular and the ephemeral" [28.10, p. xiv]. Early Twitter studies which identified the majority of tweets to be "pointless babble" [28.33, p. 5] or "daily chatter" [28.124, p. 62] casted doubt on the value of Twitter as a meaningful communication medium [28.10]. This reduction to banal content has led many in academia to consider tweeting a waste of time and to therefore reject Twitter as a means of scholarly communication [28.125]. In particular, the 140-character limit of tweets has many scholars doubting Twitter's usefulness for research. This may be why Twitter is one of the best known and at the same time least used social media platforms in the scholarly community [28.13, 30–32, 125].

The adoption of new technologies is often met with functional and psychological barriers. Lack of time and skills and negative perceptions of platforms have been identified as barriers to using social media in academia [28.126]. Reluctance often stems from the notion that tweeting wastes precious time and introduces challenges that come with the blurred boundaries of professional and personal communication. This mix of professional and personal identities on Twitter specifically has been revealed by many studies and has been identified as a potential reason not to use Twitter in academia. Even though in the general public, Twitter uptake is higher among young adults, it is often early career researchers who are more reluctant to tweet about their work [28.127, 128]. Young academics feel the highest pressure to publish in high-impact journals and limit their time spent on social media, and feel more vulnerable when publicly exposing their ideas, particularly to uncertain audiences [28.128–131]. An academic interviewed about the future of scholarly communication expressed their concern regarding the use of Facebook and Twitter for work purposes [28.129, p. 97]:

> There's this research group in my area and, for some reason, they're really into Facebook. So they want to do a lot of discussions on Facebook and that type of thing. But I really just don't have time. It's like Twittering. I just can't ... There needs to be a little bit of space where I can actually think about something. And I think for some people, they're just wired in such a way that they like that constancy, and they are also able to actually say something intelligent quickly. And I'm not like that. I have to be a little bit more deliberate and think about things a little bit more. And so I can't Twitter ... I need some time to reason...

However, early-career researchers were often more likely to find social media useful in the context of scholarly communication and collaboration [28.36, 38, 132]. *Bowman* [28.41] identified a u-shaped relationship between Twitter use and academic experience: US professors seven to nine years into their academic careers were more likely to use Twitter compared to those with fewer and more years of experience. A humanities scholar addressed the issue of blurred boundaries on Twitter [28.127, p. 59]:

> I think it can be distracting, especially to grad students, when they're trying to navigate, when they're needing to learn, adopt, and use these new technologies, but at the same time learn to discriminate among technologies that are more for social things but are being used in the name of research. The lines are too blurry.

The tension that arises between scholars who take to Twitter and those who are reluctant to discuss scholarly matters on social media is demonstrated by an incident where a genomics paper published in *PNAS* was criticized on Twitter for flaws in study design and analysis casting doubt on its conclusions. The tweet sent by a genetics researcher at the University of Chicago included charts and tables from a reanalysis of the *PNAS* paper's data, which he published in the open access and open-peer review journal *F1000Research*. The tweet and reanalysis provoked many responses on Twitter and in the comment section of the *F1000Research* paper, some of which demanded the retraction of the *PNAS* paper [28.89], while the *PNAS* authors accused the critic of violating the norms of science by taking to Twitter.

The clashing of personal and professional and the fuzzy boundaries between the two has also lead to severe negative consequences for scholars. Tweets by faculty have caused outrages among students, other faculty members, university administration and the public at large. Identifying professionally on Twitter has affected the academic careers of some scholars and, in a certain case, controversial tweets have provoked death threads [28.133]. Tweeting had serious effects on the career of a tenured University of New Mexico psychology professor, who had fat-shamed students on Twitter: "Dear obese PhD applicants: if you didn't have the willpower to stop eating carbs, you won't have the willpower to do a dissertation #truth." The professor was asked to apologize and had to undergo sensitivity training, while his work was monitored and he was banned from working on the graduate students admission committee for the rest of his career [28.96, 134, 135].

In another incident, an English professor lost his tenure-track position offer from the University of Illinois at Urbana-Champaign due to a tweet that was interpreted as anti-Semitic [28.94]. In response to withdrawal of the job offer by the University of Illinois, the American Association University Professors updated their report on academic freedom to reinstate that academic freedom applies to comments made on social media [28.93].

28.3 Scholarly Output on Twitter

As this chapter discusses scholarly Twitter metrics, it focuses on how scientific papers are diffused via tweets. Although, as described above, only a small amount of scholars' tweeting activity involves linking to publications [28.14, 71], tweets to journal articles represents one of—if not the—most popular altmetric. This might be due to the significance of publications in peer-reviewed journals in scholarly communication as well as the ease at which tweets that mention or link to document identifiers (e. g., the Digital Object Identifier DOI) can be retrieved. Another reason that Twitter-based altmetrics—and, in fact, all altmetrics—gravitate towards journal articles is that they aim to complement existing bibliometric measures, which reduce scholarly output in a similar manner.

In the following, the altmetrics literature is reviewed to provide an overview of currently used scholarly Twitter metrics for journal articles. As the majority of available studies only scratch the surface of what can potentially be extracted from Twitter activity, the literature review is complemented by an analysis of tweets collected by Altmetric.com. This analysis goes beyond tweet counts and correlations with citations and aims to reflect the *What*, *How*, *When*, *Where* and *Who* of scholarly publications shared on Twitter. This includes what types of documents are tweeted, how Twitter affordances such as hashtags, retweets and @mentions are used to share them, when and where articles are tweeted and who is tweeting them. Although this chapter is not able to provide solutions to problems associated with creating meaningful metrics from tweets, particular limitations and pitfalls are highlighted, while demonstrating the potential of available data. Together with the findings on Twitter use described above, the chapter tries to provide context to aid in the interpretation of different metrics. It thus aims at improving the understanding of what Twitter-based scholarly metrics can and cannot reflect.

28.3.1 Data and Indicators

The analysis of tweets that mention scholarly documents is based on Twitter data collected by Altmetric.com until June 2016, which contains 24.3 million tweets mentioning 3.9 million unique documents. Altmetric started systematically collecting online mentions of scholarly publications in 2012 and is a particularly valuable data source for tracking Twitter activity related to scholarly output, as it continuously stores tweets that mention scholarly publications with a DOI. Through accessing tweets through the Twitter API firehose, Altmetric circumvents the usual issues of Twitter data collection that researchers are confronted with when using the freely available Twitter APIs. While the Twitter *Streaming* API limits access to a random sample of 1% of tweets, the *Representational State Transfer* (REST) API is rate-limited and the *Search* API restricts access to only the most recent tweets relevant to a particular query [28.136]. As Altmetric started data collection in 2012, Twitter activity is incomplete for documents published earlier.

Altmetric's Twitter data is matched to bibliographic information from WoS using the DOI. This match between document metadata and tweets affords the possibility to determine the amount of scholarly output that does and does not get tweeted. The link to WoS data also provides access to cleaner and extended metadata of tweeted documents, such as the publication year, journal, authors and their affiliations, and a classification system of scientific disciplines. At the same time, the match of the two databases also excludes tweets to publications not indexed in WoS and thus comes with the known restriction and biases of WoS coverage. This is why the following analysis describes results for two datasets, the first containing all 24.3 million tweets covered by Altmetric (dataset A), and the second the 3.9 million tweets mentioning documents with a DOI, covered by WoS and published in 2015 (dataset B). The number of unique documents in dataset A is based on the Altmetric ID. As Altmetric.com's metadata is based on multiple sources, one publication might be treated as two documents, particularly if it has multiple versions, such as a journal article on the publisher's website and a preprint on *arXiv*. Similar duplications are possible but less likely in dataset B, which is based on unique identifiers and cleaner metadata in WoS.

In the following, a set of descriptive indicators are used based on tweets to scholarly documents and associated metadata. Table 28.1 provides an overview of each metric. As described above, the focus here is on journal articles but the metrics can, nevertheless, be applied to any scholarly document or other research object such as scholarly agents including "individual scholars, research groups, departments, universities, funding organizations and others entities acting within the scholarly community" [28.119, p. 376]. Similar to most bibliometrics, the metrics described in Table 28.1 can be applied to any aggregated set of documents, such as all documents relevant to a certain topic, published in the same journal, by the same author, institution, country or in a specific language.

28.3.2 What Scholarly Output Is Tweeted?

As shown above, only a small share of tweets sent by scholars actually link to scholarly output. Tweets linking to blogs or other websites are often more frequent than those linking to scientific publications. At the same time, the great majority of altmetric studies on Twitter focus on peer-reviewed scientific journal articles with a DOI, which represent one of the main limitations of currently captured altmetrics [28.3, 137–139].

In comparison to other common altmetrics, Twitter is the platform which exhibits the second largest activity related to scientific papers, following the social reference management platform Mendeley [28.15, 140]. The disciplinary differences in Twitter uptake described above are equally visible, and likely resulting in large difference in Twitter coverage between scientific disciplines. In the majority of studies, usually between 10% to 30% of selected documents were mentioned on Twitter at least once [28.7, 8, 15, 19, 141–144]. As the majority of tweets linking to scholarly papers occurs immediately after publication [28.6, 113] and Twitter activity increased annually, Twitter coverage increases by year of publication, with the most recent papers being more likely to be tweeted and older papers hardly getting shared on Twitter [28.8]. For example, while more than half of 2015 PLOS papers were tweeted at least once [28.53], Twitter coverage was at 12%

Table 28.1 Scholarly Twitter metrics associated with scholarly documents

Type of metric	Scholarly Twitter metric	Description
Tweets	Twitter coverage	Percentage of documents with at least one tweet
	Number of tweets	Sum of total number of tweets
	Twitter density	Mean number of tweets per document
	Twitter intensity	Mean number of tweets per tweeted document
Retweets	Share of retweets	Percentage of tweets that were retweets
	Retweet density/intensity	Mean number of retweets per document/tweeted document
Users	Number of users	Unique number of users associated with a document
	User density/intensity	Mean number of users per document/tweeted document
	Mean number of followers	Mean of the number of followers of users tweeting a document
Hashtags	Hashtag coverage	Percentage of documents with at least one hashtag
	Number of hashtags	Unique number of hashtags associated with a document
	Hashtag frequency	Sum of total number of hashtag occurrences
	Share of hashtags	Percentage of tweets with at least one hashtag
	Hashtag density/intensity	Mean number of hashtags per document/tweeted document
@mentions	@mention coverage	Percentage of documents mentioning a user name
	Number of mentioned users	Unique number of users mentioned in tweets associated with a document
	@mention frequency	Sum of total number of @mentions
	@mention density/intensity	Mean number of @mentions per document/tweeted document
Time	Tweet span	Number of days between first and last tweet
	Tweet delay	Number of days between publication of a document and its first tweet
	Twitter half-life	Number of days until 50% of all tweets have appeared

for those published in 2012 [28.19]. Similarly, 13% of 2011 [28.7], 16% of 2011–2013 [28.145] and 22% of 2012 WoS documents [28.15] had received at least one tweet, while Twitter coverage increased to 36% for WoS 2015 papers in dataset B (Table 28.2).

Coverage varied between disciplines and journals, but also between databases and geographic regions. Twitter is blocked in countries like Iran and China, which reflects on the visibility of their authors and papers. For example, the share of tweeted papers was low for papers published by authors from Iran [28.146]. Such geographical biases affect Twitter visibility and need to be taken into account when comparing Twitter impact of documents, authors and institutions from different countries. For example, as few as 6% of Brazilian documents published 2013 and indexed in SciELO [28.147] and 2% of a sample of Iranian papers covered by WoS had been tweeted [28.146]. On the contrary, at 21% Twitter coverage of Swedish publications seems to be more in line with general findings [28.143]. As the sample of Iranian publications included documents published between 1997 and 2012, and 98% of tweeted papers were published between 2010 and 2012, both geographical and publication date biases influence Twitter coverage.

Particularly high coverage was found for *arXiv* submissions [28.113, 148]. This high activity was, however, not caused by high Twitter uptake in the physics, mathematics and computer science communities, but created by Twitter bots [28.149]. The extent of such automated rather than human activity and its implications for the meaning of scholarly Twitter metrics are further discussed below.

While Twitter coverage describes the extent to which a set of documents gets diffused on Twitter, Twitter density (i. e., the mean number of tweets per paper) reflects the average tweeting activity per document. In general, each paper receives less than one tweet on average, with large variations between disciplines (see below). As Twitter density is influenced by Twitter coverage and thus particularly low when only a small share of papers gets tweeted, Twitter intensity reflects the average tweeting activity for tweeted papers only, excluding non-tweeted papers [28.15]. The number of tweets per paper is usually heavily skewed, much more than citations. For example, 63% of tweeted biomedical papers were only tweeted once [28.8].

A similar distribution can be observed for the two datasets described above (Fig. 28.1). The 24.3 million tweets in dataset A mentioned 3.9 million documents (based on Altmetric ID), 43% of which were tweeted once, 19% twice, 10% three times, while only 4% of documents received more than 25 tweets. The document with the most Twitter activity was mentioned 35 135 times by 144 users, with one user tweeting 34 797 times. This highlights that the number of distinct users per document might be a better proxy of diffusion than the total number of tweets.

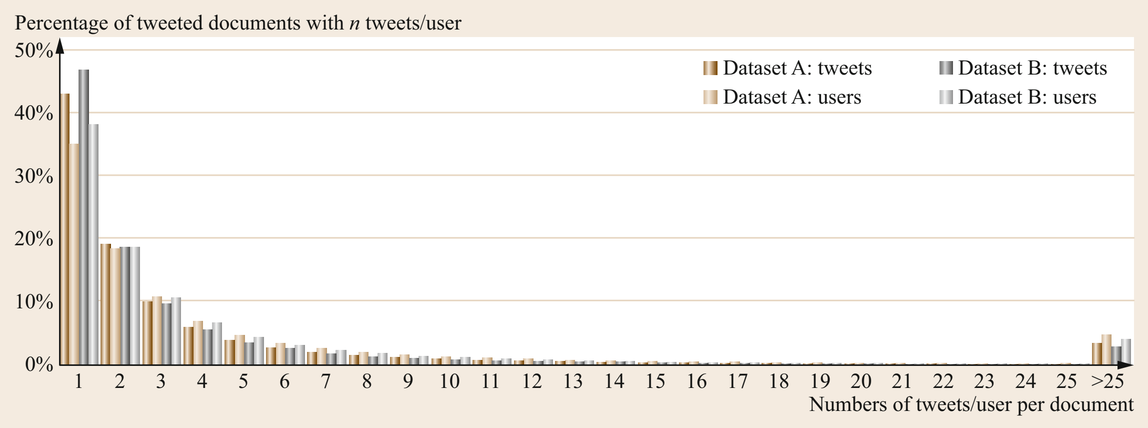


Fig. 28.1 Number of tweets and users per tweeted document for datasets A and B

Tweets in dataset A were sent by 2.6 million users and link to 3.9 million documents, amounting to a Twitter intensity of 6.2 tweets per tweeted paper. The number of users is based on unique Twitter handles included as *Author ID on Source* in Altmetric's Twitter data. Since the data does not include Twitter's unique user ID, users might be counted more than once if they changed their Twitter handle. Very likely caused by the 140-character limitation, less than 1% of tweets link to more than one document. As shown in Fig. 28.1, the distribution of number of users per document is even more skewed with 47% of all documents tweeted by a single user only. The distribution for WoS 2015 papers (dataset B) is similar but slightly less skewed for documents (based on WoS identifier) mentioned in one tweet and by one user only; 35% and 38% of documents were tweeted once or by one user, respectively. The total of 3.9 million tweets were sent by 601 290 users mentioning 548 841 documents, which corresponds to a Twitter coverage of 36%, intensity of 7.2 and a density of 2.6 tweets per document.

Disciplines and Journals

Twitter activity varies among disciplines and even journals of the same field. As there is no gatekeeping in Twitter, these variations might not be entirely due to actual impact of a particular journal but can be heavily influenced by individuals and marketing strategies of publishing houses or other stakeholders. Despite these variations, studies have shown that multidisciplinary and biomedical journals as well as social science publications are particularly visible on Twitter, while the so-called hard sciences are tweeted about less [28.8, 15, 19, 53, 140–144, 150, 151].

Previous findings are corroborated by the analysis of discipline and journal-based tweeting activity in dataset B. As shown in Table 28.2, over one third of 2015 documents in WoS have been tweeted, which represents a significant increase compared to papers published in the previous years. Twitter coverage shows large variation between disciplines, from 59% of publications in biomedical research, health and psychology, to less than 10% in mathematics and engineering & technology. This supports previous findings that Twitter activity is particularly elevated around publications from the biomedical and social sciences [28.7, 15].

It should be emphasized that Twitter activity in the arts and humanities cannot be generalized, as DOIs are not commonly used in these disciplines. While overall, 76% of all documents had a DOI, as few as 19% of all WoS-indexed journal articles in the arts were linked to this unique identifier. DOI use does not only differ between disciplines but also by country or language of publication, which is why results may be biased in these regards as wells. Considering *National Science Foundation* (NSF) specialties with more than half of its papers having a DOI, Twitter coverage was highest in parasitology (78%), allergy (76%) and tropical medicine (70%) and lowest in metals & metallurgy (1%), miscellaneous mathematics (2%), mechanical engineering (2%) and general mathematics (3%). With more than 10 tweets per document, Twitter density was highest in general & internal medicine (13.5) and miscellaneous clinical medicine (12.3).

At a Twitter coverage of 100%, 198 of 9340 tweeted journals had all of their documents diffused on Twitter, which strongly suggests a systematic and automated diffusion, possibly by a dedicated account maintained by the journal or publisher (see below). Among journals with at least 100 papers with a DOI in 2015, *JAMA* and *Biotechnology Advances* had the highest

Table 28.2 Dataset B: Twitter activity for WoS articles published in 2015

Discipline	Papers 2015		Tweeted papers		Tweets				Users	
	N	DOI coverage (%)	*N*	Twitter coverage (%)	*N*	Density	Intensity	% RTs	*N*	Intensity
All disciplines	**2 014 977**	**76**	**548 841**	**36**	**3 960 431**	**2.6**	**7.2**	**50**	**601 290**	**6.6**
Natural sciences & engineering										
Biology	124 402	73	33 945	37	226 575	2.5	6.7	54	52 235	4.3
Biomedical res.	226 011	84	112 470	59	1 025 061	5.4	9.1	50	229 851	4.5
Chemistry	158 929	90	32 829	23	81 739	0.6	2.5	34	13 860	5.9
Clinical medicine	663 481	64	223 641	52	1 784 438	4.2	8.0	51	288 226	6.2
Earth & space	96 792	91	25 616	29	136 732	1.6	5.3	51	42 641	3.2
Engr. & tech.	253 020	88	18 441	8	50 577	0.2	2.7	33	17 235	2.9
Mathematics	51 240	85	2890	7	17 441	0.4	6.0	43	7826	2.2
Physics	128 766	93	16 783	14	55 295	0.5	3.3	30	14 822	3.7
Social sciences & humanities										
Arts[a]	18 995	19	566	15	1962	0.5	3.5	39	1134	1.7
Health	51 535	74	22 662	59	191 530	5.0	8.5	52	60 090	3.2
Humanities[a]	76 998	39	4601	15	19 150	0.6	4.2	52	9784	2.0
Prof. fields	54 109	72	14 760	38	93 936	2.4	6.4	47	41 449	2.3
Psychology	40 657	78	18 735	59	145 767	4.6	7.8	51	50 974	2.9
Social sciences	70 042	76	20 902	39	135 193	2.5	6.5	53	51 288	2.6

[a] Results are too low to be representative of the publication output of the respective description.

Twitter density at 115.4 and 113.2 tweets per paper, respectively. With more than 10 000 unique Twitter users, *PLoS One*, *BMJ*, *Nature*, *Science*, *PNAS*, *NEMJ*, *JAMA*, *Lancet*, *Scientific Report*, *Nature Communications*, *JAMA Internal Medicine*, *PLOS Biology*, *British Journal of Sports Medicine*, *Cell*, *Biotechnology Advances* and *BMJ Open* were tweeted by the largest audience on Twitter (Table 28.3). These journals reflect large multidisciplinary and biomedical journals with large readership, which have been previously identified as highly tweeted, possibly because they exhibit a particular relevance to people's everyday lives [28.140]. Almost all of these journals maintain official Twitter accounts appearing among the three most tweeting users, which suggests that Twitter has become an important platform for journals and publishers to promote their contents. The number of unique Twitter users, particularly for the general science journals, seems to be low in comparison to global readership. For example, while *Nature* claimed to have 3 million unique visitors per month [28.152], as few as 42 365 Twitter users mentioned a 2015 paper. However, the numbers seem to be in line with print circulation [28.153].

Going beyond peer-reviewed journals indexed in WoS, other sources (based on Altmetric metadata) of scholarly documents are also frequently shared on Twitter. In fact, the largest number of tweeted documents in dataset A came from *arXiv*: a total of 319 411 *arXiv* submissions were tweeted 1.1 million times by 110 134 users. As shown in Fig. 28.2, the number of tweeted documents and unique users derivate particularly for the most popular sources. Although *arXiv*, *PLoS One* and *SSRN* are the most popular platforms according to the number of tweeted documents, *Nature*, *The Conversation* and *PLoS One* are tweeted by the largest number of users.

The majority of popular sources are peer-reviewed journals indexed in WoS, which also lead the ranking in dataset B (see above). Apart from these, links to repositories for documents (arXiv, SSRN, bioRxiv) or data (figshare, Dryad), as well as to websites like *The Conversation* or *ClinicalTrials.gov*, are also frequently tweeted. It should be mentioned that the document metadata in Altmetric is based on a variety of sources and thus shows some inconsistencies. For example, the source is unknown for 6% of tweeted documents and 4% of tweets in dataset A, and some sources appear in various spellings. Therefore, the number of distinct sources (49 379) or journal IDs (28 457) represents an overestimation of different sources. These inconsistencies can also be found in the publication year and other document metadata, which are essential to characterize what kind of scholarly output has been diffused on Twitter.

Table 28.3 Twitter metrics for journals with the largest Twitter audience based on number of users (≥10 000 users) including three most active users

Journal	Twitter coverage (%)	Tweets N	Density	Intensity	%RTs	Users N	Top 3 most active users (according to number of tweets; official journal accounts are marked in italic)
PLoS One	70	148 494	5.2	7.4	47	59 210	aprendedorweb, uranus_2, *PLOSONE*
BMJ	91	149 874	45.3	49.7	66	57 622	*bmj_latest*, bookapharmacist, npceorg
Nature	100	85 523	89.0	89.3	56	42 365	Doyle_Media, randomshandom, TheRichardDoyle
Science	99	78 141	70.0	70.8	67	39 225	*sciencemagazine*, PedroArtino, rkeyserling
PNAS	89	72 312	19.9	22.4	54	36 496	abbrabot, EcoEvoJournals, uranus_2
NEMJ	99	74 263	75.2	75.9	62	33 142	medicineupdate, *NEJM*, JebSource
JAMA	95	68 420	115.4	121.1	64	31 018	*JAMA_current*, robarobberlover, ehlJAMA
Lancet	98	45 752	69.9	71.6	65	25 673	medicineupdate, darmtag, *TheLancet*
Scientific Reports	56	41 177	4.6	8.3	50	23 395	*SciReports*, uranus_2, geomatlab
Nature Communications	78	39 540	11.8	15.1	52	21 274	*NatureComms*, Kochi_Study, PatrickGoymer
JAMA Internal Medicine	91	48 882	99.8	110.1	69	20 197	*JAMAInternalMed*, JAMA_current, GeriatrialNNSZ
PLOS Biology	99	24 481	90.0	91.3	70	12 878	*PLOSBiology*, PLOS, nebiogroup
BJ of Sports Medicine	97	26 121	75.7	77.7	69	11 618	*BJSM_BMJ*, exerciseworks, SportScienceNI
Cell	96	23 519	38.2	39.9	57	11 081	Brianxbio, *CellCellPress*, topbiopapers
Biotechnology Advances	55	14 823	113.2	205.9	67	10 331	robinsnewswire, GrowKudos, ElsevierBiotech
BMJ Open	76	16 882	11.5	15.1	60	10 090	*BMJ_Open*, LS_Medical, SCPHRP

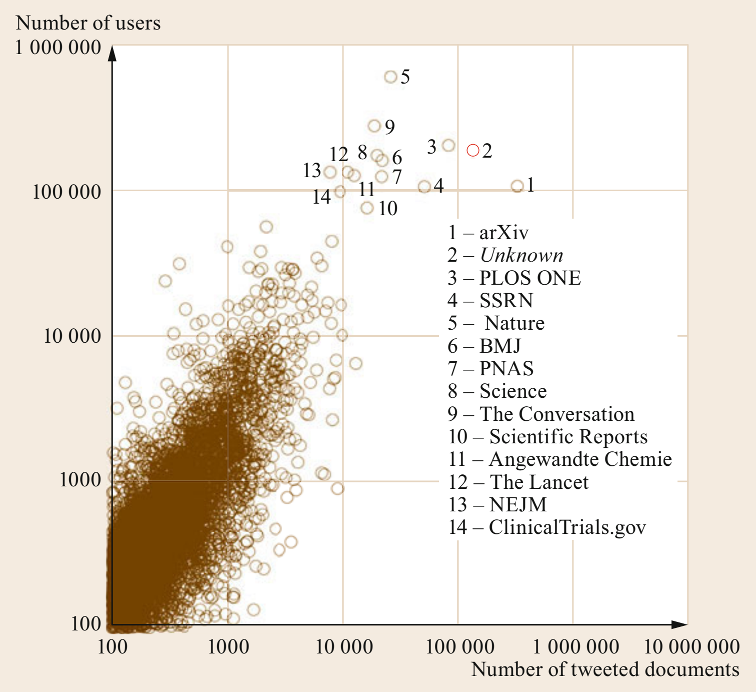


Fig. 28.2 Number of tweeted documents and unique users per source in dataset A (≥100 tweeted documents)

Document Characteristics

Besides disciplinary and topical differences, studies have also focused on determining what type of documents are popular on Twitter. Papers with particularly high Twitter visibility often had humorous or entertaining contents rather than scientific merit [28.8, 154]. A study that coded title characteristics of 200 highly tweeted papers found that 16 included a cultural reference (i. e., proverbs, idioms, fictional characters, music) and 13 were humorous or light [28.155]. *Bornmann* [28.156] reported that among papers recommended on *F1000*, those labeled as *good for teaching* were frequently tweeted. *Andersen* and *Haustein* [28.142] found that meta-analysis and systematic reviews received significantly more tweets than other medical study types.

Marking a clear distinction from citation patterns, particularly high Twitter activity was found for document types that are usually considered *uncitable* [28.15]. For example, Twitter coverage for news items was twice as high as that for all documents. Twitter density was highest for news items (3.0), editorial material (1.6) and reviews (1.4), by far exceeding the overall average of 0.8 tweets per document. The success of these document types on Twitter suggests that [28.15, p. 8]:

> documents that focus on topical subjects, debates and opinions, which are probably presented in simpler and less technical language, are more likely to appear and become popular on Twitter.

Publications with shorter titles, fewer pages and references tend to receive more tweets, while the opposite tendencies are usually observed for citations [28.15].

Similarity to Other Types of Usage

Since altmetrics were proposed as an alternative or complement to traditional bibliometric indicators, most studies analyze to what extent these new impact metrics correlate with citations. The motivation behind correlation studies lies in determining whether tweeting patterns are comparable to citing behavior. Positive correlations between citation and tweets would indicate that tweets measure something similar to but much earlier than citations, making tweet-based indicators predictors of impact on the scholarly community [28.6]. Early studies argued that low or negative correlation coefficients would indicate a different type of impact than that on citing authors, and possibly an impact on society in general.

The first study analyzing the relationship between tweets and citations found a significant association between highly tweeted and highly cited papers published in the *Journal of Medical Internet Research*, as highly tweeted publications identified 75% of those which were later highly cited [28.6]. However, this claim was based on 55 papers in a journal which itself maintained a strong Twitter presence. For a set of 4606 *arXiv* submissions, tweets were a better predictor of early citations than downloads [28.113]. The generalizability of the finding that tweets were early indicators of citation impact was later refuted by a large-scale analysis based on 1.3 million documents, which found low correlations overall between tweets and citations [28.8].

Correlations between tweets and citations vary between datasets due to particular differences between disciplines or journals, but are low overall, between 0.1 and 0.2 [28.7, 8, 15, 19, 53]. It should be noted that correlations are affected by low Twitter coverage and thus differ whether untweeted papers are excluded or included from the analysis [28.15]. Instead of replacing citations as a faster and better filter of relevant publications, as was suggested in the altmetrics manifesto [28.157], the number of tweets seem to mirror visibility on other social media platforms, in particular Facebook, rather than visibility within the community of citing authors [28.15, 53]. If nothing else, the difference between tweet and citation counts as reflected in low correlations might be due to the fundamental difference between the act of citing and tweeting [28.119].

A moderate negative correlation was found comparing publication output and tweeting activity of a group of astrophysicists on Twitter, suggesting that researchers who tweet a lot focus their efforts on communication and outreach rather than publishing peer-reviewed articles [28.110]. This inverse relationship between a researchers' standing in the scholarly community and their visibility on Twitter has led to the so-called Kardashian Index, a tongue-in-cheek indicator that reveals that those who tweet more publish less and vice versa [28.16].

Rather than correlating tweets and citations, *Allen* et al. [28.158] aimed to measure the effect of promoting articles on social media (including Twitter) on usage statistics. Comparing the number of views, downloads and citations of randomly selected articles published in PLoS One before and after promoting them on social media, article views and downloads increased significantly, but citations one year after publication and social media metrics did not.

28.3.3 How is Scholarly Output Tweeted?

Not the least due to the evaluation community's focus on counts, altmetrics research has focused much less on tweet content than on correlations and other quantitative measures. Among those looking at tweet content, the focus has been on the analysis of Twitter-specific affordance use [28.41, 63]. This includes in particular the use and analysis of hashtags, retweets and @mentions, which are further described below.

Analyzing 270 tweets linking to journal articles, *Thelwall* et al. [28.88] found that 42% contained the title of the article, 41% summarized it briefly and 7% mentioned the author. As few as 5% explicitly expressed interest in the article. While sentiment was absent in the great majority of tweets, 4% of tweets were positive and none negative. Similarly, a large-scale study that automatically identified sentiments in tweets using *SentiStrength* found that the majority of tweets were neutral and that, if sentiment was expressed, it was positive rather than negative [28.159]. Based on 192 832 tweeted WoS documents published in 2012, 11% of 487 610 tweets were positive, 7% were negative and 82% did not express any sentiment after removing the article title words from tweets. Tweets linking to chemistry papers were the least likely to express sentiments [28.159, 160].

Retweets and @mentions

As described above, retweets and @mentions represent a particular form of conversational tweets, which seemed to enjoy particular popularity among academic Twitter use. Half of the 4 million tweets linking to 2015 WoS papers were retweets (Table 28.2), which suggests that a significant amount of tweeting activity reflects information diffusion that does not involve much engagement. Compared to the studies investigating retweet use among general Twitter users [28.23] and academics [28.14, 48, 71, 110], the share of retweets among tweeted journal articles is rather high. The percentage of retweets tends to be lowest in disciplines with low Twitter coverage, which suggests that users in disciplines with low Twitter uptake do not use it as much for information diffusion, possibly because they are not as well connected. In 32 of 120 NSF specialties with DOI coverage above 50%, retweets exceed original tweets (Table 28.2): retweeting was particularly common in miscellaneous zoology, general & internal medicine, miscellaneous clinical medicine and ecology with retweet rates above 60% and low in solid-state physics, inorganic & nuclear chemistry, chemical physics and applied chemistry with less than 20%.

Hashtags

Thirty-one percent of the 24.3 million tweets captured by Altmetric until June 2016 contained a hashtag, which is comparable to other studies on hashtag use by academics [28.48, 110], but far higher than the 5% among a random sample of tweets in 2009 [28.23]. 401 287 unique hashtags were mentioned 12.6 million times, which amounts to an average occurrence of 31 per unique term. A total of 105 705 unique hashtags were used in tweets linking to 2015 WoS papers. While 33% tweets contained a hashtag, 46% of all articles were described with at least one hashtag. Each hashtag was mentioned on average 21 times for a total hashtag frequency of 2.2 million. Hashtag frequency is extremely skewed, as 3% and 6% of hashtags are responsible for 80% of hashtag occurrences in dataset A and B, respectively. For example, the most popular hashtag in dataset A was used 162 754 times (1.3% of all occurrences), while 169 992 hashtags only occurred once. Figure 28.3 demonstrates on a log-log scale the number of tweets in which hashtags were mentioned, as well as the number of distinct users mentioning each hashtag. While in general, a linear relationship can be found between the number of occurrences and users, a few popular hashtags are tweeted only by limited number of users, indicating a smaller community.

The most popular hashtags in dataset A were #science (Table 28.4, 1.3% of hashtag occurrence), #cancer (0.9%), #physics (0.8%), #openaccess, #health (0.7%), #paper, #oa and #research (0.5% each). The occurrence of #oa as well as #openaccess among the most frequent hashtags reflects the known heterogeneity of folksonomies and the need for *tag gardening* when trying to analyze topics [28.161]. WoS 2015 papers (dataset B) were most frequently tagged as #cancer (1.0%), #health, #openaccess, #science (0.9%), #FOAMed, #Diabetes, #ornithology and #Psychiatry (0.6%). The order changes when considering the number of unique users instead of tweets per hashtag. Among hashtags that occurred at least 1000 times, the largest discrepancy between the number of tweets and users can be observed for #genomeregulation (1924 tweets; 10 users), #eprompt (2281; 17 users) and #cryptocurrency (4515; 38 users), which were, on average, tweeted more than 100 times by the same users. On the contrary, the user–hashtag ratio was lowest for #Fit (4818; 4743), #StandWithPP (*Stand with Planned Parenthood*; 1060; 972), #dataviz (1010; 912), #coffee (1517; 1246), and #PWSYN (title of popular science book *The Patient Will See You Now*; 1017; 834), which indicates widespread adoption among Twitter users. Accordingly, these hashtags are more general and less scientific.

Table 28.4 shows hashtag-based stats for both datasets. As to be expected, hashtag frequency and the number of unique terms is greater for dataset A, as it covers all documents in Altmetric and the whole timespan, while dataset B is restricted to WoS documents published in 2015. On average, each hashtag occurred in 21 tweets, was used by 13 users to tag 7 documents and 4 journals indexed in WoS. As shown by the per-

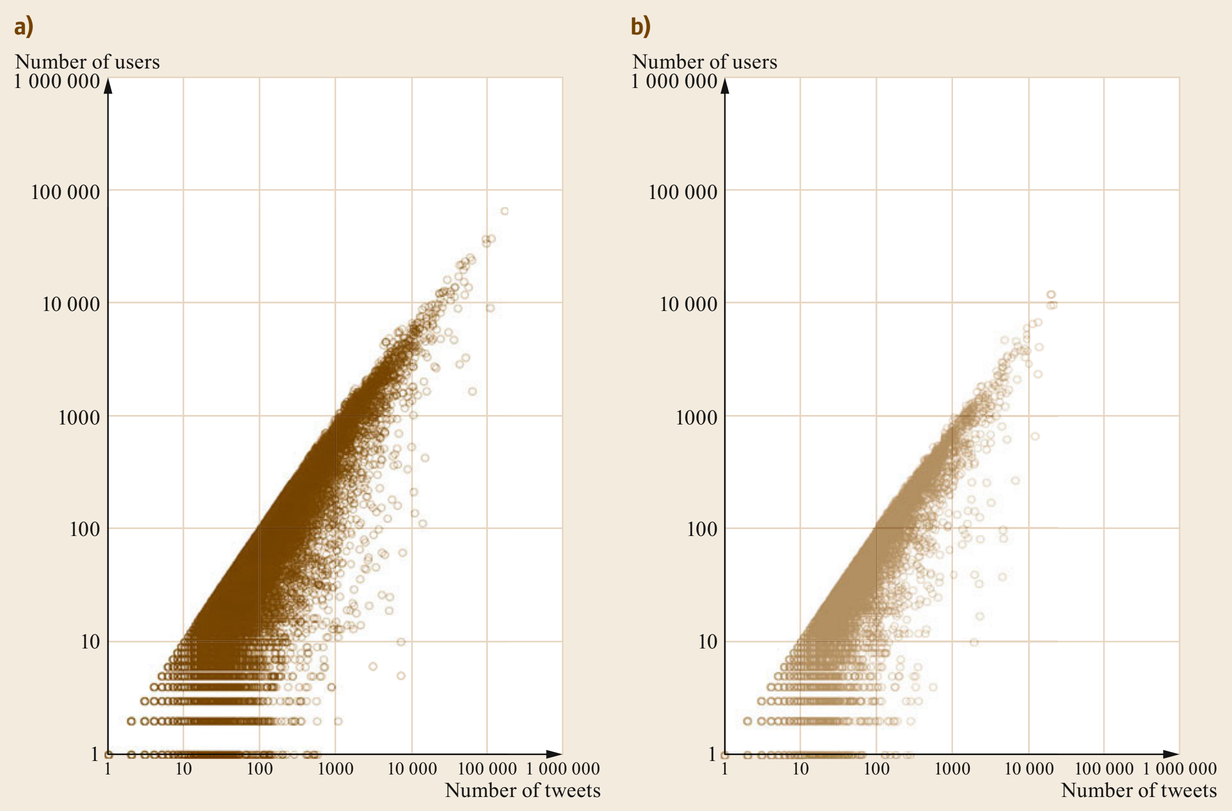


Fig. 28.3a,b Number of users and number of tweets per hashtag for all tweets captured by Altmetric **(a)** and for tweets to 2015 WoS papers **(b)**

Table 28.4 Hashtag statistics for datasets A and B with most frequent hashtags based on number of tweets

Most frequent hashtags	Statistics of hashtag frequency	Number of tweets	Number of users	Users per hashtag	Number of documents	Number of journals	Tweet span
Dataset A							
#science	Mean	31	17	1	11		233
#cancer	*Standard deviation*	*617*	*237*	*4*	*209*	n/a	*459*
#physics	Minimum	1	1	1	1		0
#openaccess	Maximum	162 754	65 334	1425	54 845		1825
#health	99th percentile	395	229	6	129		1746
#paper	90th percentile	20	15	2	8		1036
#oa	75th percentile	5	4	1	2		182
	50th percentile	2	2	1	1		0
Dataset B							
#cancer	Mean	21	13	1	7	4	102
#health	*Standard deviation*	*222*	*111*	*2*	*73*	*19*	*192*
#openaccess	Minimum	1	1	1	1	1	0
#science	Maximum	21 122	12 056	274	9765	1684	1723
#FOAMed	99th percentile	311	186	5	92	49	831
#Diabetes	90th percentile	20	15	2	7	6	381
#ornithology	75th percentile	6	5	1	2	2	130
	50th percentile	2	2	1	1	1	0

centiles, hashtag occurrence is extremely skewed. On the individual level, the number of users, documents and journals associated with a hashtag can provide information as to how general and widespread a hashtag is, or how specific and relevant to only a small group of users. The timespan, that is, the number of days between the first and last occurrence of a hashtag, indicates its topicality or timeless relevance. For example, among hashtags that occurred at least 1000 times, #diet, #water and #nutrition were used during the course of more than four years to describe 2015 documents, while #XmasBMJ lasted only 73 days. The first tweet linking to a 2015 WoS paper with the #diet hashtag appeared 17 November 2011. The discrepancy between tweet date and publication date can be described by the lag between online date and journal issue date (see [28.114] for an analysis of the publication date problematic).

28.3.4 When is Scholarly Output Tweeted?

Tweet activity related to scholarly documents has been shown to occur shortly after publication and disappear within a few days [28.6, 113]. Tweeted half-lives and delay between publication and first tweet can thus be measured in hours rather than days. This short-lived attention also points to Twitter being used to diffuse new papers instead of discussing them intensely.

It is, however, challenging to accurately calculate delay and decay for all publications in WoS as the publication date of the journal does not sufficiently represent when a publication was actually available. Even with the more accurate article-level information of online dates, there are issues to determine the actual date of publication, as demonstrated by tweets mentioning articles before they were supposed to be published [28.114]. Due to the inaccuracy of available publication dates, tweet delay and tweeted half-lives are not computed.

As shown in Fig. 28.4, there are clear differences between weekdays and weekends, reflecting patterns of the work week, which has also been shown to have an effect on journal submissions [28.162] and download patterns [28.163]. During the week, tweeting activity increases from Monday (14% of tweets) to a peak on Wednesday (18%), and decreases again towards the weekend. Twitter users tweet, on average, 23% more about scholarly documents on a Wednesday and 41% less on a Sunday. Figure 28.4 also shows the different magnitude of Twitter activity among years, as well as an overall increase throughout each year. While a general increase from January to December can be observed for each of the four years of tweets, the general trend also reflects the academic year: activity is higher in spring and fall and drops slightly in summer and particularly during the winter break during the last two and first weeks of each year. Considering that Twitter activity often climaxes the day of or day after publication, the season and weekday of publication might influence a document's visibility on Twitter [28.114]. *Zhang* and *Paxson* identified Twitter bots based on tweeting patterns that were too regular to be human [28.164].

28.3.5 Where is Scholarly Output Tweeted?

Twitter provides the possibility to geotag each tweet with precise latitude-longitude information of a user's current location. However, since this function is not activated by default, it is only rarely used. Less than 5% of tweets contain geo coordinates [28.165, 166], which is why geotags of tweets are not a reliable source to determine where scholarly output is tweeted. Another data source to determine the geographic distribution

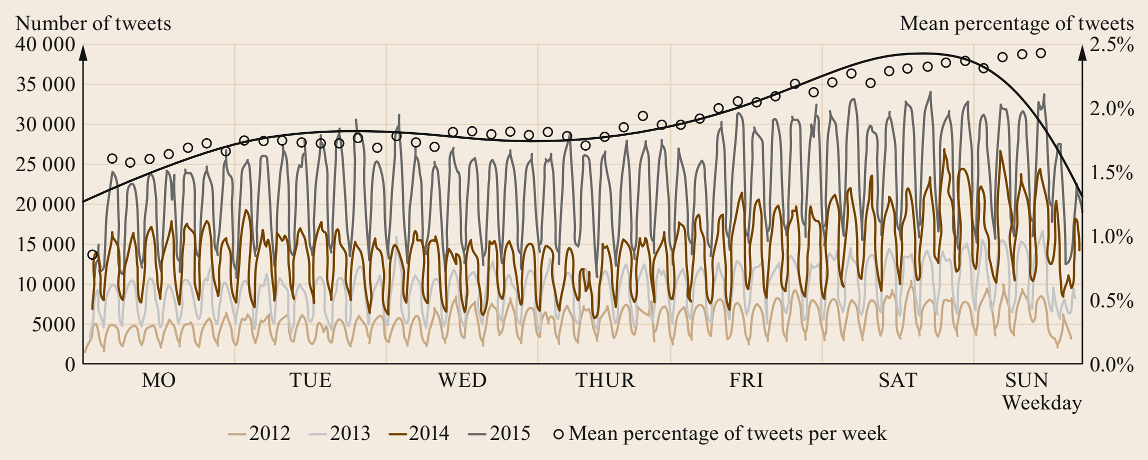


Fig. 28.4 Number of tweets per week day per year and mean of the weekly percentage per year

of Twitter users is to analyze the location information provided in the Twitter bio. However, since the profile location is usually not automatically generated but freely edited by users, it cannot be used without extensive data cleaning. *Takhteyev* et al. [28.167] showed that 8% of a sample of 3360 Twitter profiles contained specific latitude-longitude information, 57% named a location and 20% a country, while 15% used fictional places (e. g., Hogwarts) or too general descriptions to determine the users' whereabouts. Using this profile information, Altmetric is able to determine location information for two-thirds of its tweets. It becomes apparent that in many cases information is not accurate enough to determine the exact location, as remote locations in the UK and Kansas are among the most frequent tweet locations [28.168].

Due to these limitations, the analysis of where users tweet scholarly documents is restricted to the country level (Table 28.5). Altmetric provides location information for 57% of users, 58% of tweets and 71% of documents (dataset A) and 58% of both users and tweets and 70% of documents for the WoS subset (dataset B). Users from the US are overrepresented, as 20% of tweets are sent by Twitter users with an identified location in the US, followed by the UK (8%), Canada (3%), Japan, Australia and Spain (2% each). A similar distribution can be observed for the 2015 WoS articles (dataset B), as the top 10 countries by number of users stay the same, although the UK, the Netherlands, Spain and Canada, and to a lesser extent Australia and Germany, gain in percentage of users, while India, France, Japan and the US lose in comparison to dataset A. While altmetrics have been marketed as democratizers of science evaluation in terms of having the potential to correct for biases created by WoS and other US and English-centric journal databases, these results show that, when it comes to Twitter, known biases persist or are even intensified on social media.

28.3.6 Who Tweets Scholarly Output?

One of the main motivations for considering tweets as an altmetric indicator is that Twitter is used by the general public and thus, at least theoretically, offers insight into how non-academics engage with scholarly output. In order to separate tweets by the public from those sent by members of the scholarly community, Twitter users have to be identified and classified as such.

Identifying Users who Tweet Scholarly Content

One of the main challenges of determining the type of impact reflected by tweets to scientific papers is to identify who is tweeting. While Mendeley provides certain standardized user demographics such as academic status, discipline or country for users associated with a paper, the classification of tweets by user type is restricted to Twitter bios. These self-descriptions are 160-character texts, which provide users with the space to present themselves to other users of the microblogging platform.

Applying a codebook to determine who tweets scientific papers based on Twitter username, bio and photo, a sample of 2000 accounts tweeting links to articles published in *Nature*, *PLoS One*, *PNAS* and *Science*, *Tsou* et al. [28.60] found that almost one quarter of accounts were maintained by an organization.

Table 28.5 Top 10 countries by number of users for dataset A and B

	Dataset A			Dataset B			
	Documents	Tweets	Users		Documents	Tweets	Users
Number of unique items	3 903 064	24 343 105	2 622 117		548 841	3 960 431	601 290
Some country information	71%	58%	57%		70%	58%	58%
Missing country information	69%	42%	44%		77%	42%	42%
Top 10 countries by number of users							
US	33.0%	19.8%	20.1%	US	36.2%	19.5%	19.4%
UK	21.6%	11.0%	8.3%	UK	27.4%	12.3%	10.6%
CA	7.9%	2.9%	3.1%	CA	9.3%	3.1%	3.4%
JP	4.3%	1.8%	2.3%	ES	9.6%	3.3%	2.7%
AU	6.5%	2.9%	2.3%	AU	7.6%	2.5%	2.4%
ES	7.5%	2.8%	2.2%	JP	4.4%	1.5%	2.0%
FR	5.3%	1.5%	1.4%	NL	4.0%	1.0%	1.3%
IN	2.6%	0.7%	1.1%	FR	7.9%	1.6%	1.2%
NL	3.5%	1.0%	1.1%	DE	5.6%	1.1%	0.9%
DE	4.1%	1.0%	0.9%	IN	3.4%	0.7%	0.9%

Among these were mainly non-profits (42%), corporations (29%) and universities (13%), while many were also classified as news, media or outreach institutions (19%). Among the 1520 accounts identified as individuals, two thirds were male. One third of the users were identified as having a PhD and 12% as students. This amounts to almost half of all identified individuals having completed or pursuing a doctorate degree, which stands in glaring contrast to about 1% of the US population with a PhD [28.60], strongly suggesting that it is the scholarly community rather than the general public who tweets links to scientific papers.

Applying a similar codebook to a random sample of 800 accounts tweeting 2012 WoS papers, 68% of accounts were maintained by an individual, 21% by an organization, while 12% could not be identified [28.169], corroborating the findings by *Tsou* et al. [28.60]. Among individuals, 47% used professional terms (e. g., doctor, MD, photographer) to describe themselves, 22% identified as researchers (e. g., scientist, professor, postdoc), 13% as science communicators (e. g., writer, author, journalist, blogger) and 7% as students (e. g., grad student, PhD candidate). Reflecting the blurred boundaries between personal and professional communication, many individuals used words from more than one of these categories to describe themselves. For example, 8% of accounts were classified as researchers and professionals and 5% as professionals and science communicators [28.169]. Science communicators were also the largest group of Twitter users mentioned by astrophysicists [28.50]. Although labor-intensive and based on little more than 160-character self-descriptions, the above studies show that it is feasible to extract members of academia from users tweeting scholarly documents. Keyword-based searches can be applied to identify scholars in larger samples, but are limited by either low recall or low precision depending on the particular query [28.53].

It is considerably more challenging to identify members of the general public. Although many Twitter bios contain terms depicting personal lives (e. g., father, wife, yoga lover), the presence of these terms does not necessarily mean that accounts are maintained by non-academics, because scholars often describe themselves in both a personal and professional manner on Twitter [28.41, 48, 59]. Similarly, it is challenging to distinguish members of the public based on an indeterminate list of terms of non-academic professions (e. g., consultant, photographer), especially when also considering accounts in languages other than English. Even the comparably straight-forward identification of a researcher who strictly identifies as such on Twitter, becomes problematic when they shared a paper out of private interests. For example, a tweet by a physicist might actually reflect engagement by the public rather than scholarly communication, if they tweeted about a cancer study as a member of a patient group rather than in their academic role.

An alternative to classifying users based on publicly available information is to approach them directly and ask them who they are. *Alperin* [28.57] pioneered such a survey method on Twitter which, with the help of an automated Twitter account, asked users who had tweeted a *Scielo Brazil* paper whether they were affiliated with a university. Such a direct approach might also be helpful for determining the motivation for a user to tweet a specific paper, helping to quantify and distinguish different types of tweets, such as endorsement or critical discussion, diffusion or self-promotion. Author self-citations or self-tweets accounted for 7% of a sample of 270 tweets [28.88].

The 24.3 million tweets captured by Altmetric were sent by 2.6 million users. Looking at the most active users who tweeted more than 1000 times during the whole period covered by Altmetric (Table 28.6), the presence of accounts automatically diffusing scholarly articles on Twitter becomes apparent [28.149]. In fact, 15 of the 19 most productive accounts in Table 28.6 with more than 25 000 tweets self-identified as bots (see below).

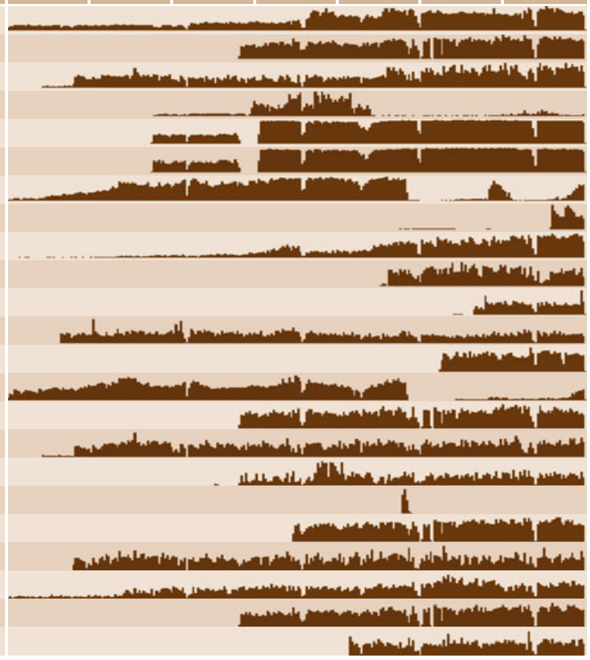
Classifying Users by Twitter Activity

Instead of classifying users according to their self-descriptions, accounts can also be grouped based on their activity. Dividing Twitter accounts into three groups of top 1%, 9% and 90% of users (according to number of tweets) helps to distinguish lead and highly active users from less active ones [28.170]. This classification provides insights into tweeting behavior of different types of users. Separating the 601 290 users in dataset B by number of tweets linking to a 2015 WoS article, 6016 lead users, 54 535 highly active and 540 739 least active users can be identified. Lead users contributed between 84 and 19 973 tweets, with a median of 149 tweets per user, had on average 935 followers (median = 935; mean = 3862) and tweeted the 2015 papers during an average tweet span of 598 days. Highly active users contributed between 9 and 83 tweets (median = 16), had less followers (median = 442.5; mean = 2136) and shorter tweet spans (mean = 388 days), while least active users tweeted up to 8 times (median = 1), had 212 followers and were active for a period of 58 days.

Lead (top 1% of users), highly active (9%) and least active (90%) users contributed 43%, 31% and 25% of tweets to the entire set of 2015 WoS papers, respectively (Fig. 28.5). Interestingly, these percentages differ among NSF disciplines with least active users overrep-

Table 28.6 Number of followers, tweets, tweet span, tweets per day and tweeting activity per week for the most active users in dataset A ($\geq$20 000 tweets per user)

Twitter handle	Number of followers	Tweets	Tweet span	Tweets per day	Tweeting activity per day						
					MON	TUE	WED	THUR	FRI	SAT	SUN
blackphysicists	12 914	92 583	1826	50.7							
MathPaper	1889	73 239	1086	67.4							
anestesiaf	1428	63 953	1713	37.3							
oceanologia	389	62 585	1353	46.3							
UIST_Papers20XX	35	50 211	1360	36.9							
UIST_Papers19XX	6	49 838	1359	36.7							
hiv_insight	14 328	48 714	1822	26.7							
russfeed	2127	44 463	577	77.1							
uranus_2	2519	42 053	1790	23.5							
Immunol_papers	477	40 721	639	63.7							
psych2evidence	513	40 658	410	99.2							
InorganicNews	1044	36 869	1657	22.3							
arXiv_trend	166	33 272	457	72.8							
hlth_literacy	5582	32 337	1823	17.7							
AstroPHYPapers	3427	31 154	1086	28.7							
cirugiaf	406	30 692	1710	17.9							
ThihaSwe_dr	601	29 357	1164	25.2							
semantic_bot	0	28 908	28	1032.4							
CondensedPapers	530	27 603	912	30.3							
libroazuln	128	24 994	1613	15.5							
rnomics	1520	24 905	1821	13.7							
PhysicsPaper	476	22 894	1086	21.1							
epigen_papers	788	21 916	739	29.7							



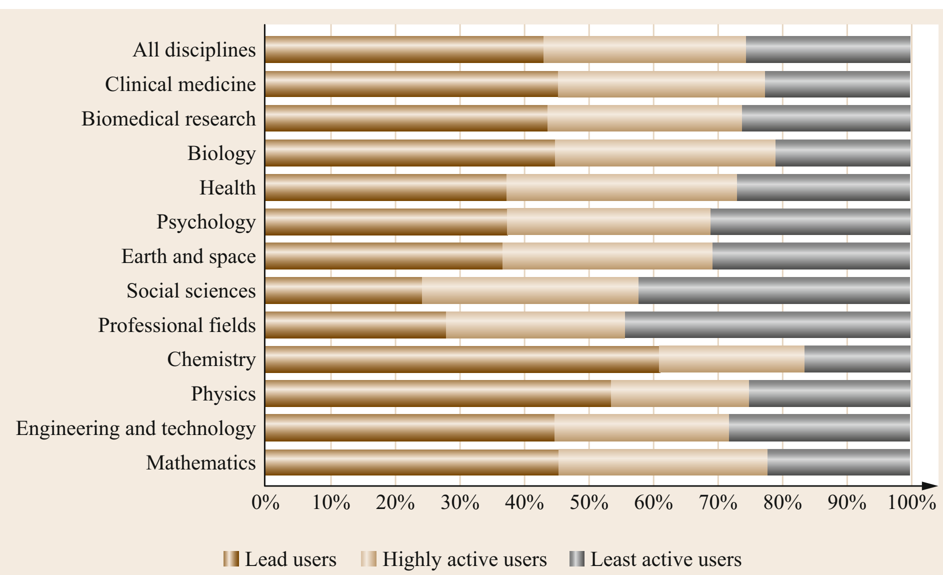


Fig. 28.5 Percentage of tweets from lead users (1%), highly active users (9%) and least active users (90%) per discipline (dataset B)

resented among those tweeting literature from professional fields, social sciences, psychology and earth and space. On the contrary, lead users were overrepresented in chemistry, physics, mathematics and engineering & technology, which were the fields exhibiting the lowest Twitter coverage, density and number of unique users (Table 28.2). Assuming that the general public is least active when it comes to tweeting about scholarly papers, they are more likely to engage with articles published in journals from the professional fields and social sciences and less likely to tweet chemical papers. The high presence of lead users in chemistry and physics might, at least partly, be caused by accounts promoting these papers automatically, such as @blackphysicists and @MathPaper, which were the two most active accounts in dataset A, tweeting 51 and 67 scholarly documents per day (Table 28.6).

Twitter Bots

Automated Twitter accounts have become prevalent on Twitter [28.171]. About one fifth of tweets sent during the 2016 presidential election were estimated to be sent by bots [28.172] and almost one quarter of tweets in 2009 came from accounts tweeting more than 150 times per day. *Ferrara* et al. [28.173, p. 96] define social bots as "a computer algorithm that automatically produces and interacts with humans on social media, trying to emulate and possibly alter their behavior". Automated Twitter accounts can be further distinguished between useful bots and antisocial or spambots [28.171, 173].

Bots are also infiltrating academic Twitter. Among a random sample of 800 Twitter accounts captured by Altmetric, 8% seemed completely and 5% partially automated [28.169], while automated accounts who self-identified as such were responsible for 9% of tweets to *arXiv* submissions [28.149]. *Shuai* et al. [28.113] even removed half of the tweets to a sample of tweets to documents on *arXiv*, as they were created by bots. Regardless of whether Twitter accounts that automatically tweet scientific papers are considered useful or spam, it is safe to say that their automated tweets do not reflect *impact*. In the context of altmetrics and tweets to scientific papers, bot activity thus needs to be at least identified, if not entirely removed in an impact assessment. Although spammers and excessive self-promotion was identified as a challenge by Altmetric, they still considered gaming a rare and easy-to-identify threat in 2013 [28.174].

Twitter's terms of service specifically address automation. While prohibiting spam, Twitter encourages automated tweets if they "broadcast helpful information" [28.175]. However, what is considered spam continuously evolves as users apply "new tricks and tactics" [28.176] to adapt to or circumvent Twitter rules. Twitter bots can be identified based on specific regularities in their tweeting behavior, such as the frequent and repetitive use of the same hashtags, URLs, tweet format and content, and regular temporal activity, as well as the follower/friends ratio, @mentions to non-followers or account suspensions [28.171]. Sixteen percent of Twitter accounts were identified as automated based on not-uniform-enough or too-uniform tweeting patterns to be stemming from a human [28.164]. Social network indicators based on the Twitter follower and friend network are considered more robust measures, as they are harder to influence [28.177]. The *BotOrNot* algorithm additionally considers linguistic features and tweet sentiment to detect automated Twitter accounts [28.178]. However, as these spamming measures get known and integrated by Twitter to block spam accounts, bots employ more sophisticated algorithms to avoid getting caught, resulting in an arms race between those who create and those who seek to identify Twitter spam.

Analyzing the most active users who have tweeted at least 1000 times (dataset A), a keyword-based query searching the Twitter bio as well as user name and handle revealed that among 2043 accounts, 248 identified themselves as automated (Fig. 28.6) and 305 as journal or publisher accounts (Fig. 28.7). These make up 30% and 11% of the tweets sent by the 2043 most active users (Table 28.7), which correspond to 7% and 3% of the entire 24.3 million tweets to scholarly documents in dataset A. The median number of followers is significantly lower than other accounts (Table 28.7) and as few as 6% of the 1.8 million tweets sent by the 248 accounts contain @mentions. If scholarly bots do mention other users, they often seem to reference journals such as @hiv_insight, which frequently mentions @PLoSMedicine, @STI_BMJ and @JAMA_current. Other than social bots in the general Twittersphere, scholarly bots seem to not try to emulate human behavior or game the system. They rather resemble RSS feeds tweeting the paper title and a link, often specifying what type of information they diffuse. Some even provide instructions to create similar feeds. For example, the Twitter bio of @asthma_papers reads "RSS feed for #asthma papers in #Pubmed. Create a feed of your own using instructions here: https://github.com/roblanf/phypapers".

(*robot* AND NOT *robotics*) OR (*bot* AND NOT (*botany* OR *robotics*)) OR (*paper* OR *publication* OR *Lit* OR *preprint* OR *article* OR *peer-review* OR *journal*) AND *feed*) OR *news feed* OR *datafeed* OR *RSS* OR *new submissions* OR (*new* AND *paper*) OR *Latest publication* OR *new publication* OR *arxiv* OR (*PubMed* AND NOT *Chief Editor*) OR *bioRxiv* OR (*papers* AND (*auto* OR *stream* OR *tweet* OR *updates* OR *links*)) OR *publication alert* OR *daily updates*

Fig. 28.6 Query used to identify automated Twitter accounts

(*journal* AND NOT (*journals* OR *journalism*)) OR *jrnl* OR *publisher* OR *publishing* OR *university press* OR *Oxford journals* OR *Cell Press* OR *CellPress* OR *Dove Press* OR *Taylor & Francis* OR *CRC Press* OR *PortlandPress* OR *Routledge* OR *Springer* OR *Elsevier* OR *Wiley* OR *Walters Kluwer* OR *SAGE Publishing*

Fig. 28.7 6 Query used to identify publisher accounts

Table 28.7 Twitter metrics for self-identified bots, journal and publisher accounts and other accounts based on users with at least 1000 tweets (dataset A)

Most active Twitter accounts (≥1000 tweets, $n = 2043$)		Followers	Tweets	Tweets per day	Tweet span
Self-identified bots	Median	212	3479	5.1	845
$n = 248$	Mean	1014	7339	14.4	923
30% of tweets	*std dev*	2781	10 390	*67.0*	*477*
	Min	0	1001	0.6	28
	Max	25 003	73 239	1032.4	1823
Journal and	Median	3199	1670	1.2	1647
publisher accounts	Mean	21 475	2249	1.7	1484
$n = 305$	*Std dev*	116 124	1874	*1.6*	*369*
11% of tweets	Min	3	1001	0.6	122
	Max	1 448 649	19 256	14.7	1822
Other accounts	Median	1535	1599	1.4	1388
$n = 1490$	Mean	5236	2408	2.9	1278
59% of tweets	*Std dev*	14 091	3670	*9.3*	*496*
	Min	0	1000	0.6	8
	Max	228 224	92 583	297.3	1826

Considering that scholars use Twitter to diffuse information and stay aware of relevant literature, automated accounts might be considered useful. However, bots have shown to be harmful to society, when they are used to influence public opinion and behavior such as political opinions or elections [28.172, 179–181] or manipulation of the stock market [28.173]. If Twitter impact became part of the scholarly reward system, Twitter bots might be able to similarly influence opinions or shape outcomes of certain research metrics.

28.4 Conclusion and Outlook

This chapter provided an overview of the use of Twitter in scholarly communication. By demonstrating *who* uses Twitter in academia and for what reasons, *what* types of scholarly outputs are diffused *how*, *where* and *when*, it aimed to add context and help to interpret any scholarly metrics derived from this and similar types of social media activity.

Research evaluators and managers were particularly excited at the prospect of an easily accessible data source that would be able to capture traces of the societal impact of research. However, perhaps unsurprisingly, the majority of tweets stem from stakeholders in academia rather than from members of the general public, which indicates that the majority of tweets to scientific papers are more likely to reflect scholarly communication rather than societal impact. At the same time, Twitter uptake in academia lacks behind Twitter user by the general public. Twitter activity is influenced by geographical and disciplinary biases and publication date. Known biases towards US and UK sources persist, rather than democratizing scholarly communication and the reward system of science.

The majority of tweets linking to scientific articles appear shortly after their publication; tweeting half-lives can be measured in hours rather than days. Moreover, one can observe weekday as well as seasonal patterns, with Twitter activity peaking Wednesdays and in the fall, and plummeting during the weekends and holiday season. Journal and publisher accounts, along with Twitter bots, contribute significantly to tweeting activity linked to academic papers, which suggests that a significant extent of tweeting activity serves promotional purposes or is automated, and reflects neither societal nor scholarly impact. A large share of tweets contains hashtags and mention either the title or a short summary of the paper they referred to. Half of all articles linking to 2015 WoS papers were retweets and the majority contained no sentiments. These tweeting characteristics emphasize particular low engagement of users linking to journal articles. The main motivation for researchers to use Twitter is information diffusion, networking and to stay up-to-date with the literature. However, the sheer brevity of tweets makes intense discussions the exception rather than the rule on Twitter.

As citation behavior and motivations for citing or not citing certain sources are biased and influenced by many factors other than a paper's significance, not every citation represents impact. However, each scholarly author is bound by scholarly norms to participate in the citation process. In some rare cases, where scholars tweet corrections to publications or journals provide tweetable abstracts and organize journal clubs, Twitter has started to be integrated in or even replace certain functions of formal journal publishing. However, in most fields, tweeting does not yet play an important role in scholarly communication. In most disciplines, Twitter uptake is low and the platform is only used in a passive or infrequent manner, or tweets reflect only a part of informal scholarly communication, such as conference chatter. Moreover, Twitter uptake varies between disciplines, countries, journals and individuals, and can be easily influenced and manipulated. The presence of automated Twitter accounts which promote certain contents becomes particularly problematic when tweet counts become the basis for measures of impact.

This is not to say that Twitter should be completely disregarded as a data source for scholarly metrics. Rather, the microblogging platform should be approached critically in terms of what kind of use and user populations it captures. By reviewing the role of Twitter in scholarly communication and analyzing tweets linking to scholarly documents in depth and beyond crude counts, this chapter attempted to provide the basis for more sophisticated and well-balanced approaches to scholarly Twitter metrics.

References

28.1 L. Bornmann: Scientific revolution in scientometrics: The broadening of impact from citation to societal. In: *Theories of Informetrics and Scholarly Communication*, ed. by C.R. Sugimoto (de Gruyter Mouton, Berlin 2016)

28.2 B. Cronin: Metrics à la mode, J. Am. Soc. Inf. Sci. Technol. **64**(6), 1091–1091 (2013)

28.3 S. Haustein: Grand challenges in altmetrics: Heterogeneity, data quality and dependencies, Scientometrics **108**(1), 413–423 (2016)

28.4 J. Priem: Altmetrics. In: *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Performance*, ed. by B. Cronin, C.R. Sugimoto (MIT Press, Cambridge 2014) pp. 263–287

28.5 C.R. Sugimoto, S. Work, V. Larivière, S. Haustein: Scholarly use of social media and altmetrics: A review of the literature, J. Assoc. Inf. Sci. Technol. **68**(9), 2037–2062 (2017)

28.6 G. Eysenbach: Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact, J. Med. Int. Res. **13**(4), e123–e123 (2011)

28.7 R. Costas, Z. Zahedi, P. Wouters: Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective, J. Assoc. Inf. Sci. Technol. **66**(10), 2003–2019 (2015)

28.8 S. Haustein, I. Peters, C.R. Sugimoto, M. Thelwall, V. Larivière: Tweeting biomedicine: an analysis of tweets and citations in the biomedical literature, J. Assoc. Inf. Sci. Technol. **65**(4), 656–669 (2014)

28.9 G. Miller: Social scientists wade into the tweet stream, Science **333**(6051), 1814–1815 (2011)

28.10 R. Rogers: Debanalising twitter. The transformation of an object of study. In: *Twitter and Society*, ed. by K. Weller, A. Bruns, J. Burgess, M. Mahrt, C. Puschmann (Peter Lang, New York 2014) pp. ix–xxvi

28.11 K. Weller, A. Bruns, J. Burgess, M. Mahrt, C. Puschmann: Twitter and society: An introduction. In: *Twitter and Society*, ed. by K. Weller, A. Bruns, J. Burgess, M. Mahrt, C. Puschmann (Peter Lang, New York 2014) pp. xxiv–xxxviii

28.12 I. Rowlands, D. Nicholas, B. Russell, N. Canty, A. Watkinson: Social media use in the research workflow, Learn. Publ. **24**(3), 183–195 (2011)

28.13 R. Van Noorden: Online collaboration: Scientists and the social network, Nature **512**(7513), 126–129 (2014)

28.14 J. Priem, K.L. Costello: How and why scholars cite on Twitter, Proc. Am. Soc. Inf. Sci. Technol. **47**(1), 1–4 (2010)

28.15 S. Haustein, R. Costas, V. Larivière: Characterizing social media metrics of scholarly papers: The effect of document properties and collaboration patterns, PLoS One **10**(3), e0120495 (2015)

28.16 N. Hall: The Kardashian index: A measure of discrepant social media profile for scientists, Genome Biol. **15**(7), 424–424 (2014)

28.17 N. Desrochers, A. Paul-Hus, S. Haustein, R. Costas, P. Mongeon, A. Quan-Haase, T.D. Bowman, J. Pecoskie, A. Tsou, V. Larivière: Authorship, citations, acknowledgments and visibility in social media: Symbolic capital in the multifaceted reward system of science, Soc. Sc. Inf. (2018), https://doi.org/10.1177/0539018417752089

28.18 B. Cronin: The evolving indicator space (iSpace), J. Am. Soc. Inf. Sci. Technol. **64**(8), 1523–1525 (2013)

28.19 J. Priem, H. Piwowar, B.M. Hemminger: Altmetrics in the wild: Using social media to explore scholarly impact, arXiv Print (2012), 1–17, https://arxiv.org/abs/1203.4745

28.20 M. Duggan, N. B. Ellison, C. Lampe, A. Lenhart, M. Madden: Demographics of Key Social Networking Platforms (2015), http://www.pewinternet.org/2015/01/09/demographics-of-key-social-networking-platforms-2/

28.21 C. Puschmann: (Micro)Blogging science? Notes on potentials and constraints of new forms of scholarly communication. In: *Opening Science*, ed. by S. Bartling, S. Friesike (Springer, Cham 2014) pp. 89–106, https://doi.org/10.1007/978-3-319-00026-8

28.22 F.J. Grajales, S. Sheps, K. Ho, H. Novak-Lauscher, G. Eysenbach: Social media: A review and tutorial of applications in medicine and health care, J. Med. Internet Res. **16**(2), e13–e13 (2014)

28.23 D. Boyd, S. Golder, G. Lotan: Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. In: *43rd Hawaii Int. Conf. Syst. Sci. IEEE* (2010) pp. 1–10, https://doi.org/10.1109/HICSS.2010.412

28.24 A. Bruns, H. Moe: Structural layers of communication on Twitter. In: *Twitter and Society*, ed. by K. Weller, A. Bruns, J. Burgess, M. Mahrt, C. Puschmann (Peter Lang, New York 2014) pp. 15–28

28.25 G. Veletsianos: Higher education scholars' participation and practices on Twitter, J. Comput. Assist. Learn. **28**(4), 336–349 (2012)

28.26 M. Nentwich: Das Web 2.0 in der wissenschaftlichen Praxis. In: *Digitale Wissenschaftskommunikation–Formate und ihre Nutzung*, *Giessen* (2011) pp. 35–53, http://geb.uni-giessen.de/geb/volltexte/2011/8227/pdf/DigitaleWissenschaftskommunikation_2011.pdf

28.27 A. Osterrieder: The value and use of social media as communication tool in the plant sciences, Plant Methods **9**(1), 26–26 (2013)

28.28 N. Pearce, M. Weller, E. Scanlon, M. Ashleigh: Digital scholarship considered: How new technologies could transform academic work, In Educ. **16**(1), 33–44 (2010)

28.29 D. Zhao, M.B. Rosson: How and why people Twitter: The role that micro-blogging plays in informal communication at work. In: *Proc. ACM 2009 Int. Conf. Support. Group Work, GROUP '09* (ACM Press, New York 2009), https://doi.org/10.1145/1531674.1531710

28.30 J. Carpenter, L. Wetheridge, S. Tanner, N. Smith: *Researchers of Tomorrow: The Research Behaviour of Generation Y Doctoral Students* (2012), http://www.jisc.ac.uk/publications/reports/2012/researchers-of-tomorrow

28.31 F. Gu, G. Widén-Wulff: Scholarly communication and possible changes in the context of social media, Electron. Libr. **29**(6), 762–776 (2011)

28.32 D. Pscheida, S. Albrecht, S. Herbst, C. Minet, T. Köhler: *Nutzung von Social Media und onlinebasierten Anwendungen in der Wissenschaft* (ZBW – Deutsche Zentralbibliothek für Wirtschaftswissenschaften – Leibniz-Informationszentrum Wirtschaft, Kiel 2013), http://www.qucosa.de/fileadmin/data/qucosa/documents/13296/Science20_Datenreport_2013_PDF_A.pdf

28.33 R. Kelly: *Twitter Study*, White Paper (Pear Analytics, San Antonio 2009), http://pearanalytics.com/wp-content/uploads/2012/12/Twitter-Study-August-2009.pdf

28.34 V. Miller: New media, networking and phatic culture, Converg. Int. J. Res. New Media Technol. **14**(4), 387–400 (2008)

28.35 K. Weller, R. Dornstädter, R. Freimanis, R. N. Klein, M. Perez: Social software in academia: Three studies on users' acceptance of Web 2.0 Services. (2010) http://journal.webscience.org/360/2/websci10_submission_62.pdf

28.36 D. Grande, S.E. Gollust, M. Pany, J. Seymour, A. Goss, A. Kilaru, Z. Meisel: Translating research for health policy: Researchers' perceptions and use of social media, Health Aff. **33**(7), 1278–1285 (2014)

28.37 M.W. Wilson, S. Starkweather: Web presence of academic geographers: A generational divide?, Prof. Geogr. **66**(1), 73–81 (2014)

28.38 C. Tenopir, R. Volentine, D.W. King: Social media and scholarly reading, Online Inf. Rev. **37**(2), 193–216 (2013)

28.39 R.N. Procter, R. Williams, J. Stewart, M. Poschen, H. Snee, A. Voss, M. Asgari-Targhi: *If You Build It, Will They Come? How Researchers Perceive and Use Web 2.0* (Research Network Information, London 2010), http://wrap.warwick.ac.uk/56246

28.40 D. Ponte, J. Simon: Scholarly Communication 2.0: Exploring Researchers' Opinions on Web 2.0 for Scientific Knowledge Creation, Evaluation and Dissemination, Ser. Rev. **37**(3), 149–156 (2011)

28.41 T.D. Bowman: *Investigating the use of Affordances and Framing Techniques by Scholars to Manage Personal and Professional Impressions on Twitter*, Dissertation (Indiana Univ., Bloomington 2015)

28.42 S. Loeb, C.E. Bayne, C. Frey, B.J. Davies, T.D. Averch, H.H. Woo, B. Stork, M.R. Cooperberg, S.E. Eggener: Use of social media in urology: Data from the American Urological Association (AUA), BJU International **113**(6), 993–998 (2014)

28.43 S. Haustein, I. Peters, J. Bar-Ilan, J. Priem, H. Shema, J. Terliesner: Coverage and adoption of altmetrics sources in the bibliometric community, Scientometrics **101**(2), 1145–1163 (2014)

28.44 E.S. Darling, D. Shiffman, I. Côté, J.A. Drew: The role of Twitter in the life cycle of a scientific publication, Ideas Ecol. Evol. **6**, 32–43 (2013)

28.45 T. Desai, A. Shariff, A. Shariff, M. Kats, X. Fang, C. Christiano, M. Ferris: Tweeting the Meeting: An in-depth analysis of Twitter activity at Kidney Week 2011, PLoS One **7**(7), 1–9 (2012)

28.46 T. Desai, M. Patwardhan, H. Coore: Factors that contribute to social media influence within an Internal Medicine Twitter learning community, F1000Research **3**, 120–120 (2014)

28.47 J. Priem, K. Costello, T. Dzuba: Prevalence and use of Twitter among scholars. In: *Metrics 2011 Symp. Informetr. Scientometr. Res* (2012), https://doi.org/10.6084/m9.figshare.104629

28.48 S. Work, S. Haustein, T.D. Bowman, V. Larivière: *Social Media in Scholarly Communication. A Review of the Literature and Empirical Analysis of Twitter Use by SSHRC Doctoral Award Recipients. Study commissioned by the Social Sciences and Humanities Research Council* (2015)

28.49 K.C. Chretien, J. Azar, T. Kind: Physicians on Twitter, JAMA **305**(6), 566–568 (2011)

28.50 K. Holmberg, T.D. Bowman, S. Haustein, I. Peters: Astrophysicists' conversational connections on twitter, PLoS One **9**(8), e106086–e106086 (2014)

28.51 K. Holmberg, M. Thelwall: Disciplinary differences in Twitter scholarly communication, Scientometrics **101**(2), 1027–1042 (2014)

28.52 J. Bar-Ilan, S. Haustein, I. Peters, J. Priem, H. Shema, J. Terliesner: Beyond citations: Scholars' visibility on the social Web. In: *Proc. 17th Int. Conf. Sci. Technol. Indic. Montreal Canada 58 Sept 2012*, Vol. 52900 (2012) pp. 98–109, https://arxiv.org/abs/1205.5611

28.53 S. Barthel, S. Tönnies, B. Köhncke, P. Siehndel, W.-T. Balke: What does Twitter measure? Influence of diverse user groups in altmetrics. In: *Proc. 15th ACM/IEEE-CE Jt. Conf. Dig. Libr., Knoxville* (ACM, New York 2015) pp. 119–128, https://doi.org/10.1145/2756406.2756913

28.54 Q. Ke, Y.-Y. Ahn, C.R. Sugimoto: A systematic identification and analysis of scientists on Twitter, PLoS One **12**(4), e0175368 (2017)

28.55 N.K. Sharma, S. Ghosh, F. Benevenuto, N. Ganguly, K. Gummadi: Inferring who-is-who in the Twitter social network. In: *Proc. 2012 ACM Workshop Workshop Online Soc. Netw. WOSN '12* (ACM Press, New York 2012) p. 55, https://doi.org/10.1145/2342549.2342563

28.56 A.T. Hadgu, R. Jäschke: *Identifying and analyzing researchers on twitter* (ACM, New York 2014) pp. 23–32, https://doi.org/10.1145/2615569.2615676

28.57 J.P. Alperin: Moving beyond counts: A method for surveying Twitter users. In: *Altmetrics15: 5 Years In, What Do We Know?* (2015), http://altmetrics.org/wp-content/uploads/2015/09/altmetrics15_paper_3.pdf

28.58 S. Haustein, T.D. Bowman, R. Costas: "Communities of attention" around scientific publications: Who is tweeting about scientific papers? In: *Soc. Media Soc. 2015 Int. Conf., Toronto* (2015)

28.59 S. Haustein, R. Costas: Identifying Twitter audiences: Who is tweeting about scientific papers? In: *Metrics 2015 ASIS&T SIG/MET Workshop, St. Louis at the Arch* (2015), https://www.asist.org/SIG/SIGMET/wp-content/uploads/2015/10/sigmet2015_paper_11.pdf

28.60 A. Tsou, T.D. Bowman, A. Ghazinejad, C.R. Sugimoto: Who tweets about science? In: *Proc. 2015 Int. Soc. Scientometr. Informetr., Istanbul* (2015) pp. 95–100, https://pdfs.semanticscholar.org/81fe/8b63188cf25648a7c592bc6b5457fee3c101.pdf?_ga=1.184338726.1264550827.1478885332

28.61 L. Bonetta: Should you be tweeting?, Cell **139**(3), 452–453 (2009)

28.62 A. Sopan, P.J. Rey, B. Butler, B. Shneiderman: Monitoring academic conferences: Real-time visualization and retrospective analysis of backchannel conversations. In: *2012 Int. Conf. Soc. Inf., Washington DC* (2012) pp. 62–69, https://doi.org/10.1109/SocialInformatics.2012.20

28.63 K. Weller, E. Dröge, C. Puschmann: Citation analysis in twitter: Approaches for defining and measuring information flows within Tweets during scientific conferences. In: *Proc. ESWC2011 Workshop "Making Sense of Microposts": Big Things Come in Small Packages*, Vol. 718 (Heraklion, Heraklion 2011) pp. 1–12, http://ceur-ws.org/Vol-718/msm2011_proceedings.pdf

28.64 B.A. Chaudhry, L.M. Glode, M. Gillman, R.S. Miller: Trends in Twitter use by physicians at the American society of clinical oncology annual meeting, 2010 and 2011, J. Oncol. Pract. **8**(3), 173–178 (2012)

28.65 D.R. McKendrick, G.P. Cumming, A.J. Lee: Increased use of twitter at a medical conference: A report and a review of the educational opportunities, J. Med. Internet Res. **14**(6), e176 (2012)

28.66 R. Mishori, B. Levy, B. Donvan: Twitter use at a Family Medicine Conference: Analyzing #STFM13, Fam. Med. **46**(8), 608–614 (2014)

28.67 A. Cochran, L.S. Kao, N.J. Gusani, J.W. Suliburk, B.C. Nwomeh: Use of Twitter to document the 2013 Academic Surgical Congress, J. Surg. Res. **190**(1), 36–40 (2014)

28.68 C. Ferguson, S.C. Inglis, P.J. Newton, P.J.S. Cripps, P.S. Macdonald, P.M. Davidson: Social media: A tool to spread information: A case study analysis of Twitter conversation at the Cardiac Society of Australia & New Zealand 61st Annual Scientific Meeting 2013, Collegian **21**(2), 89–93 (2014)

28.69 C.M. Hawkins, R. Duszak, J.V. Rawson: Social media in radiology: Early trends in Twitter microblogging at radiology's largest international meeting, JACR **11**(4), 387–390 (2014)

28.70 A. Jalali, T. Wood: Analyzing online impact of canadian conference of medical education through tweets, Educ. Med. J. **5**(3), e110–e111 (2013)

28.71 J. Letierce, A. Passant, J. Breslin, S. Decker: Understanding how Twitter is used to spread scientific messages (2010) http://journal.webscience.org/314/

28.72 W. Reinhardt, M. Ebner, G. Beham, C. Costa: How people are using Twitter during conferences. In: *Creativity and Innovation Competencies on the Web, Proc. 5th EduMedia Conf., Salzburg*, ed. by V. Hornung-Prähauser, M. Luckmann (2009) pp. 145–156, http://lamp.tu-graz.ac.at/~i203/ebner/publication/09_edumedia.pdf

28.73 J.-H. Schmidt: Twitter and the rise of personal publics. In: *Twitter and Society*, ed. by K. Weller, A. Bruns, J. Burgess, M. Mahrt, C. Puschmann (Peter Lang, New York 2014) pp. 3–12

28.74 M. Baym: The perils and pleasures of tweeting with fans. In: *Twitter and Society*, ed. by K. Weller, A. Bruns, J. Burgess, M. Mahrt, C. Puschmann (Peter Lang, New York 2014)

28.75 S. Harrington: Tweeting about the telly: Live TV, audiences and social media. In: *Twitter and Society*, ed. by K. Weller, A. Bruns, J. Burgess, M. Mahrt, C. Puschmann (Peter Lang, New York 2014) pp. 237–247

28.76 E.Y.L. Leung, D. Siassakos, K.S. Khan: Journal Club via social media: Authors take note of the impact of #BlueJC, BJOG Int. J. Obstet. Gynaecol. **122**(8), 1042–1044 (2015)

28.77 N. Mehta, T. Flickinger: The times they are a-changin': Academia, social media and the JGIM Twitter Journal Club, J. Gen. Int. Med. **29**(10), 1317–1318 (2014)

28.78 I.A. Thangasamy, M. Leveridge, B.J. Davies, A. Finelli, B. Stork, H.H. Woo: International Urology Journal Club via Twitter: 12-Month Experience, Eur. Urol. **66**(1), 112–117 (2014)

28.79 J.M. Topf, S. Hiremath: Social media, medicine and the modern journal club, Int. Rev. Psychiatr. **27**(2), 147–154 (2015)

28.80 T. Whitburn, C. Walshe, K.E. Sleeman: International palliative care journal club on Twitter: Experience so far, BMJ Support. Palliat. Care **5**(1), 120 (2015)

28.81 M.N. Kamel Boulos, P.F. Anderson: Preliminary survey of leading general medicine journals' use of Facebook and Twitter, J. Can. Health Libr. Assoc. **33**(2), 38–47 (2012)

28.82 G.J. Nason, F. O'Kelly, M.E. Kelly, N. Phelan, R.P. Manecksha, N. Lawrentschuk, D.G. Murphy: The emerging use of Twitter by urological journals: Use of Twitter by urological journals, BJU International **115**(3), 486–490 (2015)

28.83 R. Micieli, J.A. Micieli: Twitter as a tool for ophthalmologists, Can. J. Ophthalmol. **47**(5), 410–413 (2012)

28.84 T. Kortelainen, M. Katvala: "Everything is plentiful–Except attention". Attention data of scientific journals on social web tools, J. Informetr. **6**(4), 661–668 (2012)

28.85 M. Amir, B.P. Sampson, D. Endly, J.M. Tamai, J. Henley, A.C. Brewer, J.H. Dunn, C.A. Dunnick, R.P. Dellavalle: Social networking sites: Emerging and essential tools for communication in dermatology, JAMA Dermatology **150**(1), 56–60 (2014)

28.86 M. Zedda, A. Barbaro: Adoption of web 2.0 tools among STM publishers. How social are scientific journals?, Journal EAHIL **11**(1), 9–12 (2015)

28.87 N. Friedrich, T.D. Bowman, W.G. Stock, S. Haustein: Adapting sentiment analysis for tweets linking to scientific papers. In: *Proc. 15th Int. Soc. Scientometr. Informetr. Conf., Istanbul* (2015) pp. 107–108

28.88 M. Thelwall, A. Tsou, S. Weingart, K. Holmberg, S. Haustein: Tweeting links to academic articles, Cybermetr. Int. J. Scientometr. Informetr. Bibliometr. **17**(1), 1–8 (2013)

28.89 C. Woolston: Potential flaws in genomics paper scrutinized on Twitter, Nature **7553**(521), 397 (2015)

28.90 Nature Publishing Group: *Author Insights 2015 survey*. Figshare (2015), https://figshare.com/articles/Author_Insights_2015_survey/1425362

28.91 P.O. Seglen: Citations and journal impact factors: Questionable indicators of research quality, Allergy **52**(11), 1050–1056 (1997)

28.92 R. Van Noorden: Brazilian citation scheme outed, Nature **500**(7464), 510–511 (2013)

28.93 Association of American University Professors: *Academic Freedom and Electronic Communications* (2013), retrieved from https://www.aaup.org/file/Academic%20Freedom%20%26%20Electronic%20Communications.pdf

28.94 B. Herman: Steven Salaita Twitter scandal: University offers settlement, but free speech questions linger, International Business Times (2014), http://www.ibtimes.com/steven-salaita-twitter-scandal-university-offers-settlement-free-speech-questions-linger-1678854

28.95 L.A. Pasquini, N. Evangelopoulos: Organizational identity, meaning, and values: analysis of social media guideline and policy documents. In: *Proc. 2015 Int. Conf. Soc. Media Soc. SMSociety '15* (ACM, New York 2015) pp. 1–5, https://doi.org/10.1145/2789187.2789198

28.96 J. Pomerantz, C. Hank, C.R. Sugimoto: The state of social media policies in higher education, PLoS One **10**(5), e0127485 (2015)

28.97 R. Kimmons, G. Veletsianos, S. Woodward: Institutional uses of Twitter in U.S. higher education, Innov. High. Educ. **42**(2), 97–111 (2017)

28.98 D.L. Linvill, S.E. McGee, L.K. Hicks: Colleges' and universities' use of Twitter: A content analysis, Pub. Relat. Rev. **38**(4), 636–638 (2012)

28.99 N.G. Barnes, A.M. Lescault: *College Presidents Out-Blog and Out-Tweet Corporate CEO's* (UMass Center for Marketing Research 2013), http://www.umassd.edu/cmr/socialmediaresearch/collegepresidentsoutblog/

28.100 G. Greenwood: Examining the presence of social media on university web sites, J. Coll. Admiss. **216**, 24–28 (2012)

28.101 V. Prabhu, A.B. Rosenkrantz: Enriched audience engagement through Twitter: Should more academic radiology departments seize the opportunity?, J. Am. Coll. Radiol. **12**(7), 756–759 (2015)

28.102 P.S. Cahn, E.J. Benjamin, C.W. Shanahan: "Uncrunching" time: medical schools' use of social media for faculty development, Med. Educ. Online **1**, 1–6 (2013)

28.103 A. Forkosh-Baruch, A. Hershkovitz: A case study of Israeli higher-education institutes sharing scholarly information with the community via social networks, Internet High. Educ. **15**(1), 58–68 (2012)

28.104 F. Boateng, Y. Quan Liu: Web 2.0 applications' usage and trends in top US academic libraries, Library Hi Tech **32**(1), 120–138 (2014)

28.105 A. Hussain: Adoption of Web 2.0 in library associations in the presence of social media, Program **49**(2), 151–169 (2015)

28.106 J. Shulman, J. Yep, D. Tomé: Leveraging the power of a Twitter network for library promotion, J. Acad. Librar. **41**(2), 178–185 (2015)

28.107 S.K.-W. Chu, H.S. Du: Social networking tools for academic libraries, J. Librar. Inf. Sci. **45**(1), 64–75 (2013)

28.108 N. Verishagen, C. Hank: Are there birds in the library? The extent of Twitter adoption and use by Canadian academic libraries, First Monday (2015), https://doi.org/10.5210/fm.v19i11.4945

28.109 C. Wilkinson, E. Weitkamp: A case study in serendipity: Environmental researchers use of

Part D | 28

traditional and social media for dissemination, PLoS One **8**(12), e84339–e84339 (2013)
28.110 S. Haustein, T.D. Bowman, K. Holmberg, I. Peters, V. Larivière: Astrophysicists on Twitter: An in-depth analysis of tweeting and scientific publication behavior, Aslib J. Inf. Manag. **66**(3), 279–296 (2014)
28.111 Y. Mou: Presenting professorship on social media: From content and strategy to evaluation, Chinese J. Commun. **7**(4), 389–408 (2014)
28.112 T.D. Bowman: Differences in personal and professional tweets of scholars, Aslib J. Inf. Manag. **67**(3), 356–371 (2015)
28.113 X. Shuai, A. Pepe, J. Bollen: How the scientific community reacts to newly submitted preprints: Article downloads, Twitter mentions, and citations, PLoS One **7**(11), e47523–e47523 (2012)
28.114 S. Haustein, T.D. Bowman, R. Costas: When is an article actually published? An analysis of online availability, publication, and indexation dates. In: *Proc. 15th Int. Soc. Scientometr. Informetr. Conf., Istanbul* (2015) pp. 1170–1179, https://arxiv.org/abs/1505.00796
28.115 K. Weller, C. Puschmann: Twitter for scientific communication: How can citations/references be identified and measured? In: *Proc. ACM WebSci'11, Koblenz* (2011)
28.116 J. Paßmann, T. Boeschoten, M.T. Schäfer: The gift of the gab. Retweet cartels and gift economies on Twitter. In: *Twitter and Society*, ed. by K. Weller, A. Bruns, J. Burgess, M. Mahrt, C. Puschmann (Peter Lang, New York 2014) pp. 331–344
28.117 C. Warzel: Meet The Man Behind Twitter's Most Infamous Phrase, BuzzFeed News (2014), https://www.buzzfeed.com/charliewarzel/meet-the-man-behind-twitters-most-infamous-phrase?utm_term=.jwWokRoZw#.fnElNylG4
28.118 K. Holmberg: The impact of retweeting, Figshare (2014), https://doi.org/10.6084/m9.figshare.1037555
28.119 S. Haustein, T.D. Bowman, R. Costas: Interpreting "altmetrics": Viewing acts on social media through the lens of citation and social theories. In: *Theories of Informetrics and Scholarly Communication*, ed. by C.R. Sugimoto (de Gruyter Mouton, Berlin 2016) pp. 372–405, https://arxiv.org/abs/1502.05701
28.120 H.D. White: Towards ego-centered citation analysis. In: *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield*, ed. by B. Cronin, H.B. Atkins (Information Today, Medford 2000) pp. 475–496
28.121 H.D. White: Authors as citers over time, J. Am. Soc. Inf. Sci. Technol. **52**(2), 87–108 (2001)
28.122 I. Lulic, I. Kovic: Analysis of emergency physicians' Twitter accounts, Emerg. Med. J. EMJ **30**(5), 371–376 (2013)
28.123 C. Honeycutt, S.C. Herring: Beyond microblogging: Conversation and collaboration via Twitter. In: *42nd Hawaii Int. Conf. Syst. Sci., IEEE* (2009) pp. 1–10, https://doi.org/10.1109/HICSS.2009.89
28.124 A. Java, X. Song, T. Finin, B. Tseng: Why we twitter: Understanding microblogging usage and communities. In: *Proc. 9th WebKDD 1st SNA-KDD Workshop Web Mining Soc. Netw. Anal. WebKDD/SNA-KDD '07* (ACM, New York 2007) pp. 56–65, https://doi.org/10.1145/1348549.1348556
28.125 A. Gerber: Online trends from the first German trend study on science communication. In: *Science and the Internet* (Düsseldorf Univ. Press, Düsseldorf 2012) pp. 13–18
28.126 H. Donelan: Social media for professional development and networking opportunities in academia, J. Furth. High. Educ. **40**(5), 706–729 (2016)
28.127 M.E. Bulger, E.T. Meyer, G. De la Flor, M. Terras, S. Wyatt, M. Jirotka, K. Eccles, C. McCarthy Madsen: Reinventing research? Information practices in the humanities, A Res. Inf. Netw. Rep. (2011), https://doi.org/10.2139/ssrn.1859267
28.128 A. Coverdale: Negotiating doctoral practices and academic identities through the adoption and use of social and participative media. In: *Proc. Eur. Conf. eLearn., Brighton*, Vol. 2011 (2011) p. 909
28.129 D. Harley, S.K. Acord, S. Earl-Novell, S. Lawrence, C.J. King: *Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines* (Univ. of California Press, Berkeley 2010), http://escholarship.org/uc/item/15x7385g
28.130 P. Mccrea: Reflections on academic blogging as a vehicle for professional development. In: *Proceedings of the European Conference on eLearning, Brighton* (2011) p. 997
28.131 G. Veletsianos: Open practices and identity: Evidence from researchers and educators' social media participation, Br. J. Educ. Technol. **44**(4), 639–651 (2013)
28.132 A. Gruzd, K. Staves, A. Wilk: Connected scholars: Examining the role of social media in research practices of faculty using the UTAUT model, Comput. Human Behav. **28**(6), 2340–2350 (2012)
28.133 S. Rothschild, B. Unglesbee: Kansas University professor receiving death threats over NRA tweet. In: *The Dispatch* (2013), http://www.shawneedispatch.com/news/2013/sep/24/kansas-university-professor-receiving-death-threat/
28.134 M. Bennett-Smith: Geoffrey Miller, Visiting NYU Professor, slammed for fat-shaming obese PhD applicants (2013), http://www.huffingtonpost.com/2013/06/04/geoffrey-miller-fat-shaming-nyu-phd_n_3385641.html, last access: 14 April 2016
28.135 L. Ingeno: Fat-Shaming in Academe, Inside Higher Ed (2013), retrieved from https://www.insidehighered.com/news/2013/06/04/outrage-over-professors-twitter-post-obese-students
28.136 D. Gaffney, C. Puschmann: Data collection on Twitter. In: *Twitter and Society*, ed. by K. Weller, A. Bruns, J. Burgess, M. Mahrt, C. Puschmann (Peter Lang, New York 2014) pp. 55–67

28.137 J.P. Alperin: Ask not what altmetrics can do for you, but what altmetrics can do for developing countries, Bull. Am. Soc. Inf. Sci. Technol. **39**(4), 18–21 (2013)

28.138 S. Haustein, C.R. Sugimoto, V. Larivière: Guest editorial: Social media in scholarly communication, Aslib J. Inf. Manag. (2015), https://doi.org/10.1108/AJIM-03-2015-0047

28.139 M. Taylor: Towards a common model of citation: Some thoughts on merging altmetrics and bibliometrics, Res. Trends (2013), 1–6 https://www.researchtrends.com/issue-35-december-2013/towards-a-common-model-of-citation-some-thoughts-on-merging-altmetrics-and-bibliometrics/

28.140 R. Costas, Z. Zahedi, P. Wouters: The thematic orientation of publications mentioned on social media: Large-scale disciplinary comparison of social media metrics with citations, Aslib J. Inf. Manag. **67**(3), 260–288 (2015)

28.141 H. Alhoori, R. Furuta: Do altmetrics follow the crowd or does the crowd follow altmetrics? In: *2014 IEEE/ACM Jt. Conf. Dig. Libr. (JCDL)* (2014) pp. 375–378, http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6970193

28.142 J.P. Andersen, S. Haustein: Influence of study type on Twitter activity for medical research papers. In: *Proc. 15th Int. Soc. Sci. Informetr. Conf., Istanbul* (2015) pp. 26–36

28.143 B. Hammarfelt: Using altmetrics for assessing research impact in the humanities, Scientometrics **101**, 1419–1430 (2014)

28.144 S. Haustein, V. Larivière, M. Thelwall, D. Amyot, I. Peters: Tweets vs. Mendeley readers: How do these two social media metrics differ?, It – Inf. Technol. **56**(5), 207–215 (2014)

28.145 N. Robinson-García, D. Torres-Salinas, Z. Zahedi, R. Costas: New data, new possibilities: Exploring the insides of Altmetric.com, Prof. Inf. **23**(4), 359–366 (2014)

28.146 A. Maleki: Networking in science tweets linking to scholarly articles of Iran. In: *SIG/MET Workshop, ASIS&T 2014 Annu. Meet., Seattle* (2014), http://www.asis.org/SIG/SIGMET/data/uploads/sigmet2014/maleki.pdf

28.147 J.P. Alperin: Geographic variation in social media metrics: an analysis of Latin American journal articles, Aslib J. Inf. Manag. **67**(3), 289–304 (2015)

28.148 S. Haustein, T.D. Bowman, B. Macaluso, C.R. Sugimoto, V. Larivière: Measuring Twitter activity of arXiv e-prints and published papers. In: *altmetrics14: Expanding Impacts and Metrics, Workshop at Web Science Conference* (2014), https://doi.org/10.6084/m9.figshare.1041514%20

28.149 S. Haustein, T.D. Bowman, K. Holmberg, A. Tsou, C.R. Sugimoto, V. Larivière: Tweets as impact indicators: Examining the implications of automated "bot" accounts on Twitter, J. Assoc. Inf. Sci. Technol. **67**(1), 232–238 (2016)

28.150 M. Fenner: What can article-level metrics do for you?, PLoS Biol. **11**(10), e1001687–e1001687 (2013)

28.151 S.R. Knight: Social media and online attention as an early measure of the impact of research in solid organ transplantation, Transplantation **98**(5), 490–496 (2014)

28.152 Publishing Group: Announcement: A new iPad app for Nature readers, Nature **492**(7428), 154–154 (2012)

28.153 Nature Publishing Group: Media Options, https://web.archive.org/web/20110513233735/http://www.nature.com/advertising/resources/pdf/2010naturemediakit.pdf?prod_code=NATURE&prod=Nature (2010)

28.154 C. Neylon: Altmetrics: What are they good for? (2014), http://blogs.plos.org/opens/2014/10/03/altmetrics-what-are-they-good-for/

28.155 F. Didegah, T.D. Bowman, S. Bowman, J. Hartley: Comparing the characteristics of highly cited titles and highly alted titles. In: *21st Int. Conf. Sci. Technol. Indic., Valencia* (2016) pp. 1190–1195, http://ocs.editorial.upv.es/index.php/STI2016/STI2016/paper/viewFile/4543/2327

28.156 L. Bornmann: Validity of altmetrics data for measuring societal impact: A study using data from Altmetric and F1000Prime, J. Informetr. **8**(4), 935–950 (2014)

28.157 J. Priem, D. Taraborelli, P. Groth, C. Neylon: Altmetrics: A manifesto (2010), http://altmetrics.org/manifesto/

28.158 H.G. Allen, T.R. Stanton, F. Di Pietro, G.L. Moseley: Social media release increases dissemination of original articles in the clinical pain sciences, PLoS One **8**(7), e68914–e68914 (2013)

28.159 N. Friedrich: *Applying Sentiment Analysis for Tweets Linking to Scientific Papers*, Bachelor's Thesis (Heinrich-Heine-Univ. Düsseldorf, Düsseldorf 2015)

28.160 N. Friedrich, T.D. Bowman, S. Haustein: Do tweets to scientific articles contain positive or negative sentiments? In: *altmetrics15: 5 Years In, What Do We Know?, Amsterdam* (2015), http://altmetrics.org/altmetrics15/friedrich/

28.161 I. Peters: *Folksonomies Indexing and Retrieval in Web 2.0* (De Gruyter/Saur, Berlin 2009)

28.162 G. Cabanac, J. Hartley: Issues of work-life balance among *JASIST* authors and editors, J. Am. Soc. Inf. Sci. Technol. **64**(10), 2182–2186 (2013)

28.163 X. Wang, S. Xu, L. Peng, Z. Wang, C. Wang, C. Zhang, X. Wang: Exploring scientists' working timetable: Do scientists often work overtime?, J. Informetr. **6**(4), 655–660 (2012)

28.164 C.M. Zhang, V. Paxson: Detecting and analyzing automated activity on Twitter. In: *Proc. 12th Int. Conf. Passive Active Meas. PAM'11* (Springer, Berlin, Heidelberg 2011) pp. 102–111, http://www.icir.org/vern/papers/pam11.autotwit.pdf

28.165 M. Graham, S.A. Hale, D. Gaffney: Where in the world are you? Geolocation and language identification in Twitter, Prof. Geogr. **66**(4), 568–578 (2014)

28.166 M. Severo, T. Giraud, H. Pecout: Twitter data for urban policy making: an analysis on four Euro-

pean cities. In: *Handbook of Twitter for Research*, ed. by C. Levallois (EMLYON, Lyon 2015)
28.167 Y. Takhteyev, A. Gruzd, B. Wellman: Geography of Twitter networks, Soc. Netw. **34**(1), 73–81 (2012), https://doi.org/10.1016/j.socnet.2011.05.006
28.168 S. Haustein, R. Costas: Determining Twitter audiences: Geolocation and number of followers. In: *altmetrics15: 5 Years In, What Do We Know?* (2015), http://altmetrics.org/altmetrics15/haustein/
28.169 S. Haustein, V. Minik, D. Brinson, E. Hayes, R. Costas, C.R. Sugimoto: Identifying Twitter user communities in the context of altmetrics. In: *3rd Altmetrics Conference 3:AM, Bucharest* (2016)
28.170 A. Bruns, S. Stieglitz: Metrics for understanding communication on Twitter. In: *Twitter and Society*, ed. by K. Weller, A. Bruns, J. Burgess, M. Mahrt, C. Puschmann (Peter Lang, New York 2014) pp. 69–82
28.171 M. Mowbray: Automated Twitter Accounts. In: *Twitter and Society*, ed. by K. Weller, A. Bruns, J. Burgess, M. Mahrt, C. Puschmann (Peter Lang, New York 2014) pp. 183–194
28.172 A. Bessi, E. Ferrara: Social bots distort the 2016 U.S. Presidential election online discussion, First Monday (2016), https://doi.org/10.5210/fm.v21i11.7090
28.173 E. Ferrara, O. Varol, C. Davis, F. Menczer, A. Flammini: The rise of social bots, Communications ACM **59**(7), 96–104 (2016)
28.174 J. Liu, E. Adie: Five challenges in altmetrics: A toolmaker' s perspective, Bull. Assoc. Inf. Sci. Technol. **39**(4), 31–34 (2013)
28.175 Twitter. (2017, April 6). Automation rules. Retrieved from https://support.twitter.com/articles/76915
28.176 Twitter: The Twitter Rules. (n.d.), retrieved from https://support.twitter.com/articles/18311
28.177 C. Yang, R.C. Harkreader, G. Gu: Die free or live hard? Empirical evaluation and new design for fighting evolving Twitter spammers. In: *Recent Advances in Intrusion Detection*, Vol. 6961, ed. by R. Sommer, D. Balzarotti, G. Maier (Springer, Berlin, Heidelberg 2011), https://doi.org/10.1007/978-3-642-23644-0_17
28.178 C.A. Davis, O. Varol, E. Ferrara, A. Flammini, F. Menczer: *BotOrNot: A System to Evaluate Social Bots* (International World Wide Web Conferences Steering Committee, Geneva 2016) pp. 273–274, https://doi.org/10.1145/2872518.2889302
28.179 P.T. Metaxas, E. Mustafaraj: Social Media and the Elections, Science **338**(6106), 472–473 (2012)
28.180 J. Ratkiewicz, M.D. Conover, M. Meiss, B. Gonçalves, A. Flammini, F. Menczer: Detecting and tracking political abuse in social media. In: *Proc. 5th Int. AAAI Conf. Weblogs Soc. Media, Barcelona* (AAAI, Menlo Park 2011) pp. 297–304
28.181 J. Ratkiewicz, M. Conover, M. Meiss, B. Gonçalves, S. Patil, A. Flammini, F. Menczer: Truthy: mapping the spread of astroturf in microblog streams. In: *Proc. 20th Int. Conf. Comp. World Wide Web WWW '11* (ACM, New York 2011) p. 249, https://doi.org/10.1145/1963192.1963301

Stefanie Haustein

School of Information Studies
University of Ottawa
Ottawa, Canada
stefanie.haustein@uottawa.ca

Stefanie Haustein is Assistant Professor at the University of Ottawa School of Information Studies, teaching research methods and evaluation, social network analysis, and knowledge organization. Her research focuses on scholarly communication, bibliometrics, altmetrics, and the role of social media in academia. Stefanie co-directs the ScholCommLab, is an associate member of the Centre interuniversitaire de recherche sur la science et la technologie (CIRST), and affiliated researcher of the Observatoire des sciences et des technologies (OST).

29. Readership Data and Research Impact

Ehsan Mohammadi, Mike Thelwall

Reading academic publications is a key scholarly activity. Scholars accessing and recording academic publications online are producing new types of readership data. These include publisher, repository, and academic social network download statistics as well as online reference manager records. This chapter discusses the use of download and reference manager data for research evaluation and library collection development. The focus is on the validity and application of readership data as an impact indicator for academic publications across different disciplines. Mendeley is particularly promising in this regard, although all data sources are not subjected to rigorous quality control and can be manipulated.

29.1 Introduction and Overview

The act of reading an academic publication is a key point at which knowledge is transferred from the author to someone else. With the prevalence of the web and social web, scholars now often read and register academic publications online, leaving electronic records of their activities. This readership data can reveal which outputs are used as well as give insights how the scientific enterprise works. It is important to exploit such information to improve research evaluation practices and to investigate how science communication is evolving. The main advantage of using readership information rather than citations is that reading occurs before citing and can, therefore, give more timely information. A second advantage is that the reading public for academic research is wider than the citing public, since it includes students, professionals, and others.

This chapter reviews research about download and readership data for academic outputs from the perspective of its value for research evaluation. It is mainly concerned with journal articles, but books and conference papers are also briefly mentioned. The chapter has extended coverage of readership information from on-

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_29

line reference managers, such as Mendeley, because of their practical value for research assessment. It also discusses publisher usage data, since *using* an article in this context often means reading it.

The usage statistics component complements the *Usage Bibliometrics as a Tool to Measure Research Activity* chapter in this Handbook (Chap. 33), which gives an extended case study of arXiv and analyzes research evaluation at a more systemic level. The current chapter also discusses usage data from the academic social network sites ResearchGate and Academica.edu, which are changing the way in which articles are discussed and shared. Previous works reviewed usage bibliometrics [29.1], readership metrics [29.2], and social media metrics [29.3]. This chapter updates the previous readership metrics chapter [29.2] with newer topics and findings. It is intended for research evaluators, scientometricians, and bibliometricians, as well as those interested in recent changes in the scholarly communication ecosystem.

This chapter opens with a discussion of reading, terminology, and research impact assessment to set the context for the ways in which readership data may be used. It continues with a discussion of online usage data from publishers and repositories, before analyzing the effect of online reference managers on the availability of readership data and discussing social networks. Finally, the chapter discusses research evaluation applications.

29.2 Reading Research: Background and Terminology

Although knowledge has traditionally been communicated in human societies orally and by imitation, the written record is a cornerstone of modern science. Academics read scholarly publications to inform themselves for their current research, for current awareness purposes, to support their teaching, or to help them fulfil a professional advisory role related to their expertise. In addition, other professionals may read journals to inform their day-to-day practice. This is the case for medical doctors, who need to be aware of the latest developments in medical practice that are relevant to their role or expertise [29.4]. For example, 73% of non-publishing Canadian physicians read journal articles [29.5], and hospital residents consider journal articles to be valuable sources of information [29.6]. Information about which academic documents are read and by whom can help the librarians that buy them, the managers and policy makers that need to evaluate the impact of the research produced, and the scholars that investigate science itself.

Many different terms have been used for reading-related data, especially within electronic contexts. The most general is perhaps *usage*, which does not imply a reason why an item was accessed but is often employed in digital contexts as an umbrella term to describe accesses of all kinds of digital resource. It is helpful to use more specific terms, when justifiable, to aid interpretation.

The more explicit term *download* refers to a local copy being taken of an electronic resource, such as a journal article, whether by a human or a robot. The terms *hit* and *view* refer to online accesses of electronic information without necessarily downloading it. For example, a digital library visitor might *view* a web page containing the title, abstract, and metadata of an article and then *download* a full-text copy of the article to their local computer. The term *full-text download* [29.7] can be used for emphasis. Although the term download usually connotes accessing a full text version of an article, some publishers provide non-downloadable full-text copies of articles, and readers may view full-text articles online without downloading them to their local computer storage [29.8].

The most specific term is *read/reader/readership*, which implies that the item accessed has been read by a human. This term can be justified for a data source if it is reasonable to believe that the items accessed will usually be read. For example, it would be reasonable to believe that a book is usually read when it is borrowed from a library. Thus, borrowing statistics could be claimed to be (non-exhaustive) reading indicators. In contrast, most accesses of documents on some websites may be from web crawlers, and so it would not be reasonable to interpret download counts as readership counts (but see the paragraph below). Moreover, human users may also systematically download journal articles if they are concerned that their access will be interrupted before they know which articles to read [29.9]. To describe download counts as a readership indicator, evidence would be needed to connect downloads to reading. Thus, if full-text downloads from a specific source are likely to be of human origin and frequently lead to reading, then it would be reasonable to refer to full-text downloads as readership data.

More specifically, and following an argument for citations [29.10], to be a readership *indicator*, download or access counts should positively correlate with reader numbers, even if not all downloads/accesses lead to reading. If this condition is fulfilled, then the counts convey information in an information theoretic sense about how often an item has been read. To give an example, a positive correlation implies that if article X has more accesses than Y, then there is a greater than 50% chance that X has been more read than Y. In general, the higher the correlation, the higher this chance is. Aggregating sets of documents also, in general, increases this chance. Thus, a moderate positive correlation between accesses and reading would give a high chance that research group A's outputs had been read more than research group B's outputs if they had been accessed more on average. In practice, the exact number of readers of any academic document is never known because of ways of accessing documents that cannot be tracked, such as from print copies of journals. Statistical correlations, therefore, need to be supported by arguments that the untracked readings are rare compared to the tracked readings, or that the tracked readings are likely to be a relatively unbiased sample of all accesses. Alternatively, if the tracked accesses can be directly shown to correlate with research impact or value, then this additional step is unnecessary.

29.3 Readership Data from Libraries

Two traditional sources of journal readership data are sales and library circulation information—either the number of libraries that hold a journal or circulation data for that journal. Both give journal-level rather than article-level evidence and are limited by the increasing share of reading that occurs online. For university libraries, the main journal readers are students, who are likely to be readers but not citers [29.11], and it is not clear whether they are also the main consumers of academic journals, or whether there are disciplinary differences in this. These students are likely to have different uses for academic journals, and librarians need to consider this factor when analyzing readership evidence to build collections [29.12].

Circulation information has been analyzed for a long time [29.13] from practical [29.14] and mathematical modeling [29.15] perspectives to support librarians [29.1]. Some examples illustrate the approaches used and give a background for more current strategies.

The earliest usage records kept by libraries were lending or reshelving statistics, and these have been used as proxies for usage or readership data for books and journals [29.16], even though they do not cover all ways in which they may be read [29.17]. This data has been shown to correlate positively with Journal Impact Factors (JIFs), at least when comparing journals from the same discipline, and journals from specialisms that are similarly served by the library [29.18, 19]. This might be due to some readers subsequently citing articles in the journals or more cited journals tending to have more useful information. For example, *Nature* and *Science* target non-citing audiences in parallel with the acknowledged high quality of their scholarly content. In contrast, other journals primarily target a professional audience (e. g., nurses, librarians, lawyers) and may be less concerned with attracting an academic readership. High correlations between online downloads and local readership information in one context give some evidence that downloads can be good indicators of readership [29.9].

Inter-library lending information and direct observations of library users can also give readership information, although the latter is too time consuming to be routinely used [29.20].

29.4 Research Impact Assessment

The publication of peer-reviewed research is a critical scholarly activity, and the analysis of scholarly publications is important for assessing the research impact of scholars or teams in most disciplines. This assessment can be qualitative, quantitative, or both. It can be combined with other sources of evidence or judgements and it can be made for formal evaluations, formative self-evaluations, or to investigate an aspect of science or science communication.

29.4.1 Peer Review

Peer review is the central evaluation mechanism for modern science [29.21]. The value of research is judged by experts from the same field because non-specialists are less able to understand the work or its contribution to scholarship. Ultimately, however, the scholarly community is collectively responsible to the governments or others that fund them, and so experts may adjust their expectations in response. For example, this may lead a community to regard work as better if it offers societal benefits.

Peer review is also at the heart of the publication system, with journal articles usually being subject to critical evaluation before a decision is made about whether to accept them, and the same is often true for monographs [29.22]. It is also central to research funding applications [29.23], perhaps as one of a checklist of attributes to assess. In some countries, peer review is used to periodically assess research to allocate block research grants. In the UK, the Higher Education Funding Council for England (HEFCE) uses panels of field experts to evaluate the quality of research from government funded universities and other academic organizations to share out funding on a merit basis [29.24].

The main disadvantage of peer review is that it is slow and expensive, consuming a substantial amount of expert time [29.25, 26]. It is also fallible because reviewers may be consciously or unconsciously biased against others' work based on personal characteristics, research competition, or the research paradigm followed [29.27]. Moreover, even expert reviewers can disagree about the merits of papers [29.28]. Nevertheless, peer review is often the most credible single source of evaluation evidence for academic research.

Although peer review primarily assesses prior work rather than predicting future performance, experts can also assess applicants' plans if these are relevant for appointments and promotions.

29.4.2 Citation Analysis

Citation-based indicators are the primary quantitative tools to evaluate research, whether on their own, to support human judgments, or to cross-check reviewers' opinions [29.29]. Citation data may come from a recognized citation index, such as the Web of Science (WoS), Elsevier's Scopus, or Google Scholar. There are also specialist citation indexes for some fields and others with a national or linguistic scope, such as the Chinese Citation Index.

The best-known citation indicator is the JIF, which estimates the average number of citations to recently-published articles in a journal. The JIF is informally used as an indicator of the quality of a journal perhaps because of its simplicity and intuitive reasonableness. Nevertheless, there are numerous problems with the accuracy of the calculations and their ability to reflect impact [29.30, 31]. There are also major problems of over-interpretation, leading to inappropriate uses [29.32, 33], such as those that ignore disciplinary differences.

The average number of citations per publication is known to vary by field and year, and so it is not reasonable to compare the average citation count between groups of publications. Field normalized indicators, such as the Mean Normalized Citation Score (MNCS) [29.34] and the Mean Normalized Log-transformed Citation Score (MNLCS) [29.35, 36] solve this problem by normalizing citation counts for the publishing field and year, so that a score of 1 always means citation impact equal to the world average.

For individual authors, the h-index has become popular [29.37], although it is biased towards senior researchers and male researchers.

All citation-based indicators suffer from several weaknesses. At a theoretical level, citations may reflect scholars acknowledging prior work that has influenced them [29.38], but they are not used systematically for this, can be influenced by spurious factors, and can be negative [29.39, 40]. Moreover, citations only reflect knowledge advancement rather than wider contributions to academia or society [29.41, 42]. For example, academic publications can be used in education, the professions [29.43], and to inform about health [29.44, 45]. The people affected in these cases may be thought of as *pure readers* in the sense of consuming academic outputs without subsequently citing them [29.46, 47]. Governments and research funders may explicitly state the need to consider non-academic impacts in their evaluations [29.48, 49], [29.50, para.1].

An important practical drawback of citation-based indicators is the considerable time that they take to accumulate. A research team might think of an idea, submit a grant proposal, get funded, carry out their research, submit a write-up to a journal, get it reviewed and accepted, and then wait for the article to be published online or in a formal journal issue. This process might take several years, and several more years would be needed before their article has attracted a reasonable number of citations from others, who had read their work and then followed the same process to conduct related research. A citation window of 3 years is sometimes recommended for research evaluations [29.51]. Thus, whilst academic evaluations often have the aim of predicting future research excellence so that it can be funded, or the promising academic can be appointed/promoted, in practice, citation-based indicators reflect performance that occurred several years in the past.

29.5 Online Access and Download Data

In response to the time delay and limited impact coverage problems of citation analysis as well as the expense of peer review, science policy-makers and managers may seek alternative indicators to reflect wider types of research impact or to give earlier evidence of impact [29.52–54].

Another important driver towards new indicators is the changing nature of scholarly communication, with increasingly diverse types of scholarly output being published online and valued even if they are rarely cited [29.55, 56]. In parallel, an increasing amount of scholarly communication takes place in public and online, leaving traces that may be used to evaluate its importance [29.57].

Given that an academic publication must be accessed and read to be valuable but that not all readers are citers, it is logical to look to access and readership data for evidence of the wider impacts of publications [29.2]. This information can also ameliorate the time delay problem of citation analyses, because an article must be read before any action can be taken based on its contents. Hence, unless the impact of an article is predicted from its metadata (e. g., from JIFs), evidence of downloads or readership gives the earliest impact evidence. Away from research evaluation, this argument has also been made for library collection development. At the level of entire journals, readership and citation may be thought of as two separate, overlapping dimensions of the impact of research and giving librarians or evaluators information about both can help them to make more informed choices [29.58, 59].

Although rarer in research evaluation contexts, the importance of readers is a natural concern for scholars, who may submit to journals partly based on the readership that they hope to gain [29.47], librarians who choose journals primarily to service their potential readers, and editors or publishers that monitor the overall audience or sales of a journal as an indicator of its value or health [29.60].

Readership data reflect something different from citations, even when only considering academic journal articles and restricting attention to academic readers. This is because there are some types of articles that are often read but rarely cited, such as a series of annual summaries of astrophysics [29.61].

29.5.1 Online Access and Download Data for Journal Usage Assessment

Statistics about local online accesses of journals are the modern way for librarians to monitor readership levels, although they are not comprehensive. Libraries keep print copies of journals, researchers and faculty may subscribe to individual serials, and educators may photocopy articles for students. In addition, article preprints may be shared from institutional and subject repositories and home pages, as well as by email and post. These uses will not normally be recorded electronically by a library or publisher but online accesses are, nevertheless, likely to give more comprehensive and timely information than library circulation data. Online usage data may be able to separate out people that browse journals from those that read articles by accessing their full text, giving more substantial information than circulation data. Local data is more relevant to libraries than generic publisher data that covers all uses, because each library serves a user community that has specific information needs, including research interests and educational provision. Moreover, local usage data seems to be more tied to reading, since robots would presumably rarely access local library copies of articles, even though humans may still systematically download them [29.9].

Online usage information originates from the log file of a web server recording accesses of journal pages or downloads of electronic copies of articles. There are many technical pitfalls with this information, including accesses by robots and repeated accesses by individuals for spamming purposes or by accident, and so online accesses do not equate with human readership. Since publishers gain extra sales if their access statistics are higher, they do not have an incentive to clean their data from spurious downloads before delivering it to libraries or others. There have been initiatives to standardize the process of cleaning the data to ensure that compliant publishers generate credible and comparable final statistics for their end users. The main initiative for this is COUNTER, which standardizes the reporting of usage information [29.62, 63].

As discussed above, usage data is inaccurate, because it is always an incomplete record of readership, and there will also be unrecorded readers. In addition, there are other important limitations that apply to some or all contexts.

For all applications, the lack of contextual information with most usage data (Mendeley and questionnaire data are exceptions) is an important restriction. Librarians may consider usage by faculty to be more valuable to the university mission than uses by students, at least on an individual level, but can rarely distinguish between the two in their data. In addition, no current major data source gives evidence about how a document was used by the reader [29.1]. This is an advantage of citations, because, in theory, the reason for a citation can be deduced from the accompanying text.

29.5.2 Online Access and Download Data for Research Evaluation

The shift to electronic publishing has led to the widespread availability of electronic access information at the level of individual articles, such as from publisher websites. This has made usage or readership data a practical source of evidence for research evaluations. In many cases, usage and readership information can be used in a similar way to citations for impact assessment, although it has different advantages and limitations. It has also not become as generally accepted as citations for this purpose. For example, the JIF is much more widely reported and recognized than any of the proposed usage-based alternatives.

For research evaluations rather than collection development purposes, statistics that are available for entire journals but not individual articles are unhelpful, although monograph circulation data can help research evaluation in the arts, humanities, and some social sciences [29.64].

Also, for research evaluation purposes, the limited nature of local access data from individual libraries for journal articles can be resolved by forming a consortium of libraries to share data [29.65] or by substituting publisher statistics. The former may be a realistic possibility for libraries that already have a common infrastructure to access electronic journals and so that data sharing can be added as an additional service rather than a completely new contribution.

Whatever the source of usage data, its principal advantage over citations for research evaluations is timeliness, because usage logically comes before the publication of citations. A second advantage is scope, because usage data includes, but does not differentiate, readers that do not cite the work. Hence, usage data may provide a timelier source of impact evidence with a wider scope. The reason why it is rarely preferred to citations is that it is much easier to manipulate, and so it is not credible enough for formal research evaluation purposes, even if from a COUNTER-compliant source. Nevertheless, it can be valuable for informal evaluations, self-evaluations, and assessments of the science system, as well as to cross-check the results of peer review or citation analysis.

Because of the accuracy limitations of usage data, it is important to assess whether it gives evidence of academic impact before it is used for article-level research evaluations. The primary strategy so far for this is to assess the extent to which article-level usage statistics correlate with citation counts. A perfect correlation cannot be expected because of the absence of pure readers from citation statistics, but a moderate or high correlation would suggest that the usage source assessed is not too affected by manipulation or fake data from robots. In contrast, a correlation close to zero would suggest that either there are many readers that have very different needs to citers or that the results have been intentionally or unintentionally manipulated.

Correlation analyses have mostly found moderate or high correlations between downloads and citations, which tends to confirm the value of usage data. A small study ($n = 153$) of article downloads in their first week found a moderate correlation (Pearson $r = 0.5$) with WoS citations 5 years later [29.66]. Similar correlations have been found for downloads of articles in the physics preprint server arXiv in the first 6 months and their citations after 2 years [29.67], for Research Papers in Economics (RePEc) preprint downloads and citations [29.68], and for PLoS PDF downloads and citations [29.69]. In contrast, a correlation of only 0.1 was found between early downloads (2 months) and later citations (25 months) for the fast, organic chemistry journal *Tetrahedron Letters*, suggesting that for this journal, early accesses represent a different type of use to citation [29.70]. For downloads within Elsevier's ScienceDirect and Scopus citation counts, the two correlate in all disciplines at the level of journals and articles; early downloads also correlate with later citations. These correlations vary in strength by discipline; they are lowest in the arts and humanities (0.2–0.3) and reach as high as 0.8 (life sciences). Despite this, the most downloaded articles tend to differ from the most cited articles for individual journals [29.71]. Confusingly, a study of Chinese journals found higher correlations between downloads and citations within the arts, humanities, and social sciences than for other academic disciplines [29.72].

At the level of journals, various download-based indicators have been defined in similar ways to the JIF, including the Usage Impact Factor [29.65] and the Download Immediacy Index [29.73]. Correlation tests have been used to help assess the value and validity of download-based indicators, with typically weaker results than at the level of individual articles [29.74] and with some negative correlations. Usage data for the Rouen University Hospital digital library had a low positive correlation with JIFs in one study [29.75], and correlations were not significantly different from 0 for JIFs and Concordia University chemistry and biochemistry journal usage data [29.11]. A comparison of JIFs with aggregate full text download data for a set of universities found low negative correlations, suggesting that journals most used by students (the main downloaders) were the least cited [29.76], see also, [29.65]. Thus, whilst download data seems to reflect a high degree of scholarly impact at the level of individual articles, when articles are aggregated into journals, scholarly impact is substantially less important, and download data may predominantly reflect educational value.

Electronic usage data can sometimes incorporate information about the origins of the users from the internet address of their computers. It is, therefore, possible to break down the readers of an article by institution and country and perhaps also organization type, if this data is made available by publishers or web server operators [29.61, 77]. This can reveal where articles and journals have had impact. This may be relevant to national funding bodies that want to demonstrate international impact or, conversely, want to make sure that the home nation rather than a competitor is benefiting from their investment.

29.5.3 Limitations of Online Usage Data

As discussed above, usage data can include false hits, whether robot accesses or downloads by people who did not intend to read the article, and articles can be accessed from multiple print and online sources [29.78]. These limitations apply unequally between journals and even between articles, so it is not fair to compare the impact of articles using any source of download data. For example, one article's publisher download count may be half that of another because it is available free online from the author, or is available in print in addition to electronic versions [29.79]. The main disadvantage of download counts from a research evaluation perspective is that they are easy to manipulate unless extensive measures are taken to protect them [29.80]. An additional disadvantage is that the data is not transparent, because publishers do not share the identities of those that accessed an article, and so authors and evaluators have no means of verifying downloads.

29.6 Readership Data from Online Reference Managers

In addition to manual methods to collect readership information, such as surveys, reader observation and reshelving information, and computerized methods, such as library, publisher, or repository download statistics, the web has made possible an additional indirect method to find whether an article has many readers: online reference managers. A reference manager is a program that records metadata about some or all the resources that a person is interested in, typically to keep track of what they have read and to automatically generate reference lists for their documents, whether they are journal articles, conference papers, books, or reports. Reference managers like EndNote, RefWorks, CiteULike, Connotea, Mendeley, Bibsonomy, and Zotero all perform this role in different ways.

If it is assumed that reference manger users tend to record articles that they have read, then the collective databases of reference managers form a large source of information about what has been read by whom. Some reference managers do not share this information but others, such as Mendeley and Bibsonomy, do, and so reference manager records are an alternative source of readership information [29.81].

At first glance, reference manager data is an unpromising source of readership evidence. Not all readers use reference managers, and so they form an incomplete readership record. No reference manager is dominant, and so if one is used as a data source, then its information will be partial even with respect to all reference manager users. Reference manager users are likely to be a biased subset of all readers, because technophobes might avoid learning a new program and people who do not write documents that include references would have little need for them.

Nevertheless, some reference managers have an advantage over download data: their information is freely available from a single source (rather than multiple publishers), they are not affected by multiple copies of articles being available (e. g., preprints in repositories), and they seem to give more definite evidence of readership than downloads, because the latter could be from a crawler or a casual user. For this reason, they can be a more realistic data source for research evaluations than download data.

Data from reference managers that are also social websites and provide an Applications Programming Interface (API), such as Mendeley, CiteULike, and Bibsonomy, fall within the scope of altmetrics [29.82]. These are indicators derived from data harvested from social web sites via APIs. The altmetrics movement has led to the creation of many new indicators. Indicator companies, such as Altmetric.com, ImpactStory.org, and Plum Analytics, systematically collect altmetric data (including from reference managers) and make it available to publishers, scholars, and institutions [29.83]. Altmetric.com, for example, attempts to provide accurate and transparent article-level indicators [29.84]. Although it includes readership data from

Mendeley, it treats this as a secondary data source, since it is not transparent (i. e., does not reveal the identity of readers). In addition, there are public initiatives to harvest and share altmetric data, such as the one from PLoS [29.85].

The promise of altmetrics is that it will deliver faster impact evidence that encapsulates wider types of impact [29.54, 86]. Within this, reference manager data fits as a fast and wider source of evidence, since reference manager users may be students [29.87] and other non-publishing article readers. Each altmetric has its own strengths and weaknesses, and potentially reflects a different type of impact. For example, tweet citations [29.88] seem to reflect attention rather than impact and are probably the fastest indicator to accrue. Reference manager data can, therefore, be thought of as an alternative to download counts as a source of readership evidence, or as a type of altmetric to be analyzed in parallel with other altmetrics.

29.6.1 Online Reference Managers: Background

Online reference managers have broadly similar functions, but each has its own software design, individual features, and user demographics. The national, disciplinary, and age composition of the adopters of each one is likely to be influenced by its age, national and disciplinary origins, and the fit of its affordances within disciplinary missions. For example, most have Western origins and did not prioritize language support, which may have alienated potential users in China, Russia, and Japan. User demographics are unknown for most, however. The descriptions below of some of the major social sites give a flavor of their differences, but their capabilities evolve over time and so may have changed now. All are online social reference managers in the sense that they manage references, are online, and allow users to create a public profile:

- Bibsonomy (www.bibsonomy.org) manages web bookmarks as well as references and incorporates social features [29.89–91]. Users typically have about 20% more references than bookmarks [29.91]. Probably like the other sites, most references are for journal articles [29.92]. Bibsonomy provides a copy of all its data free for researchers (www.kde.cs.uni-kassel.de/bibsonomy/dumps/).
- CiteULike (citeulike.org) is free, has the basic reference manager capabilities described above, and allows users to annotate references and share them with others [29.93]. It also has discussion fora and blogs [29.94]. Because of its communication and sharing capabilities it is also a type of academic social web site.
- Mendeley (Mendeley.com) is a free reference manager [29.95] that has been bought by Elsevier and offers social networking features, such as the ability to follow other users, as well as discussion groups and the ability for users to list their own publications. It, therefore, serves as an academic social network site as well as a reference manager, although its social features do not seem to be extensively used [29.96]. Mendeley offers a free and comprehensive API to access its readership data for all articles in its catalogue, so that anyone can find out how many Mendeley users have recorded any given document within their libraries. Although Mendeley does not report which users have registered each document, it gives a breakdown of user types by status (e. g., undergraduate, professor, other professional), geographic location (country), and main academic discipline [29.97].
- Zotero (www.zotero.org) is a free, open source reference manager that originated as a Firefox web browser plugin but is now available as a separate program. It has features to support group discussions and group reference sharing.

In addition to the above, RefWorks is a reference manager owned by ProQuest, and EndNote is owned by Thomson Reuters. Neither share readership data publicly at the time of writing.

29.6.2 Online Reference Managers: Coverage

Readership data from online social reference managers need to be publicly available, or at least shared with researchers, and to have many records to be useful. If a site has few users, then these are likely to be a very biased subset of readers, so the results may be misleading. For example, article readers tend to be from the same country as the authors [29.98], so any national readership biases will translate into international readership indicator biases. If most articles do not have a record of readers in the site, then its data is unlikely to be powerful enough for research evaluation purposes unless used on a large scale to compare average reader counts (e. g., using the equalized mean-based normalized proportion cited (EMNPC): [29.36, 99]). Of the online reference managers sharing data, Mendeley has the widest coverage and probably the most users. It had records for 80% of PLoS articles compared to 31% for CiteULike [29.86] and indexed more *Nature* and *Science* articles [29.100]:

- Bibsonomy: Bibsonomy has much lower coverage of physics journal articles 2004–2008 than CiteULike and probably less than 1% [29.78]. Journal articles comprise half (47%) of the items recorded, with conference papers (25%) and books (12%) also being common [29.92].

- CiteULike: Most (65%) PloS Biology articles have a record in CiteULike [29.85]. Less than 3% of physics articles 2004–2008 are in CiteULike [29.78].
- Mendeley: Virtually all (97%) articles from *Journal of the American Society for Information Science and Technology* 2001–2011 [29.101] and *PloS Biology* (95%) have a record in Mendeley [29.85]. Most (66%) PubMed articles 2010–2012 that are also in the WoS have a Mendeley record [29.102]. For Scopus medical fields, 78% of articles had at least one reader [29.103]. Another study found high coverage for WoS Medicine 2008 articles (72%) but lower (about a third) for physics, chemistry, engineering, and technology [29.104]. Less than half of recent social sciences WoS articles (44%) are in Mendeley, varying from psychology (54%) to linguistics (34%), and only 13% of humanities articles were indexed, from education (34%) to literature (4%) [29.105]. Nevertheless, 61% of Swedish humanities journal articles from 2012 were in Mendeley [29.106]. Compared to other altmetrics from Altmetric.com, Mendeley had the highest coverage (63%) of a large sample of WoS articles [29.107]. Very few books have records: only 7% of recent WoS science and medicine volumes [29.108]. Thus, whilst Mendeley has wide coverage overall and particularly for medicine, it is weak in the humanities and very weak in some disciplines and for books. This may have changed since the studies described here however.
- Zotero. No coverage information is available.

29.6.3 Online Reference Managers: Correlation with Citation Counts

When a new indicator is proposed for an aspect of research evaluation, then the logical first method to assess whether it has any value is to calculate its correlation with citation counts on the basis that a positive result would be evidence that the data was not random and related to scholarly impact in some way [29.109]. Even though a negative or zero correlation is also consistent with a new indicator reflecting a completely different type of impact, in practice, most types of impact relate to each other to some extent, and so this test is reasonable. There is extensive evidence of this type for Mendeley and a little for CiteULike. For Mendeley, readership counts correlate positively and moderately strongly with citation counts (and peer-review judgments) in most fields, with the arts being the main exception:

- CiteULike records and citations have a significant positive correlation for *Science* and *Nature* [29.100]. Usage data dominated by CiteULike have low Spearman correlations (0.1) with JIFs for physics journals [29.78]
- Mendeley records and citations have a significant positive correlation for *Science* and *Nature* [29.100], for *PLoS ONE*, *PLoS Biology*, and *PLoS Pathogens* articles [29.86] and for selected genetics and genomics articles [29.110]. Mendeley readers have a moderate overall correlation (0.5) with WoS article citations [29.107]. For PubMed articles 2010–2012 in WoS, Spearman correlations between Mendeley readers WoS citations were positive and statistically significant in all broad disciplines except the arts. They varied from 0.2 (humanities) to 0.6 (engineering and technology), with an average of 0.5 [29.102]. For WoS articles from 2008, five social science fields had Spearman correlations of 0.4–0.6, and five humanities fields had Spearman correlations of 0.4 or 0.5 [29.105]; see also [29.111]. Similar correlations were found for science and medicine fields (0.4 or 0.5) except for engineering and technology (0.3) [29.104]. Within medicine, the correlations later (and for narrower fields) rose to 0.7 [29.103]. The most systematic analysis so far checked 325 narrow Scopus fields, finding strong positive correlations in almost all [29.112]. For books, correlations between Mendeley reader counts and citations are about 0.1 [29.108]. Engineering conference papers have a very low correlation with citation counts [29.113].

A more direct source of evidence of the value of readership counts is their correlation with peer review scores. Peer review judgments are impractical to obtain for large samples of articles unless the data is a by-product of a research evaluation. For articles published in 2008 and submitted for evaluation by subject experts in the UK's Research Excellence Framework (REF) 2014, correlations between Mendeley reader counts and expert ratings in 33 of the 36 fields examined were positive, with the highest being for clinical medicine (0.4) and the lowest for music, drama, dance, and performing arts (−0.1) [29.114]. Given that these articles were selected by their authors for being high quality, the correlations are likely to substantially underestimate the underlying level of agreement between peer judgement and Mendeley reader counts, and so this is strong evidence that in most fields Mendeley reader counts reflect the quality of journal articles. A weaker corroborating piece of evidence is that UK clinical guideline references have more Mendeley readers than comparable articles do [29.115].

29.6.4 Online Reference Managers and Reading

References can be logged within a reference manager by users who have not read them [29.116] or as part of training exercises [29.117], and so it is not clear that it is reasonable to characterize reference manager data as "readership counts". The best way to find out why users enter reference data is to ask them. A survey of Mendeley users found that most (85%) added articles to cite them, but many also added articles for professional (50%) or teaching (25%) purposes. Moreover, most added articles that they had read or intended to read. Thus, Mendeley readership data clearly represents readership and a wider type of impact than scholarly impact, although mainly still within a broad academic context [29.118]. Some articles are added for educational reasons, having many readers but few citations [29.119].

Since undergraduates use reference managers, it is logical to look to readership data for evidence of educational reading. This is undermined by evidence that undergraduates and researchers tend to register similar articles [29.35].

29.6.5 Online Reference Managers: Reader Types and Demographics

Readers of research can originate from any country in the world, from any discipline, from any academic status. They can also be professionals using the information for their work or could be members of the public with an interest in a specific topic or fact. Within these groups, some read more academic research than others, and even when considering academic researchers alone, younger researchers read and cite more [29.120, 121]. Undergraduates sometimes read scientific papers but their reading is often directed by their lecturers [29.122, 123]. To interpret the type of impact reflected by readership data, it is, therefore, important to investigate the nature of people that use online reference managers. Partial information is known for Bibsonomy and Mendeley.

In terms of geography, almost half of all Bibsonomy users are from Germany [29.91], undermining its value for general impact assessment. Probably all the major services have relatively low numbers of users from China and from countries with little scientific publishing or a low level of technology use in universities.

In terms of work status, Mendeley includes substantial numbers of undergraduates and Master's students and few non-academic users. In science, it is dominated by young users: PhD students, postgraduates, and postdoctoral researchers [29.104]. In contrast, successful senior researchers seem to avoid it [29.124], and so there is an age/seniority bias.

29.6.6 Online Reference Managers: Timeliness

Mendeley readers appear about a year before citations, on average. For four library and information science (LIS) journals, the number of citations reaches the number of readers after about 7 years [29.125]. A similar pattern of initially higher readership counts than citation counts has been found for 50 fields, although the number of years needed for citations to overtake readers varies by discipline [29.126]. Early Mendeley readership counts are also better predictors of later high citation counts than are journal impact factors or citations [29.127]. All this evidence supports the conclusion that Mendeley readership counts give statistically stronger impact evidence than citation counts in the first few years after publication.

It is common for articles to have Mendeley readers as soon as they are formally published because of the prior sharing of preprints [29.128]. This makes it possible to conduct evaluations of them immediately upon publication, if these evaluations do not require the statistical power of high average readership counts.

Most importantly, the higher number of Mendeley readers than citations in the year following publication makes Mendeley reader counts correlate more strongly than citation counts with peer-review judgments of the quality of journal articles [29.114].

29.6.7 Online Reference Managers: Research Evaluation Applications

Readership data is useful for research evaluation applications where timeliness is important and there is little risk of deliberate manipulation of the data. This excludes formal exercises where those evaluated are told the data sources in advance but allows their use for more routine academic progress monitoring.

Mendeley readership counts are useful for national-level evaluations for governments to compare their progress against that of their competitors. The goal of such evaluations is to inform policy decisions or to assess the effect of recent policy changes. For the latter case in particular, timely data is essential. Mendeley readership data is preferable to citations because of its timeliness but has the limitation that it is influenced by different levels of national uptake from its users. This is a problem because of the tendency for people to read articles from their own country. It is possible to circumvent this issue with a modeling approach to measure the amount of bias in caused by the readership demographics and then correct for them, but this strategy is imperfect because it requires assumptions or information about the evolution of national uptake of the site over time [29.129].

Funding councils are also logical users of readership data. These may monitor the average impact of the research that they fund to identify which funding streams are most successful and whether the average impact of their funded research has changed over time. Web indicators can be useful for these, because the time lag of citations would delay decisions about changing ineffective funding strategies [29.130] as well as for evidence of societal research impact [29.131]. National biases in uptake and the bias towards younger users are relatively minor problems for this, and so Mendeley readership data is a better source than citations (e. g., [29.36]), although it does not reflect societal benefits or many professional uses.

One recent application harnesses readership data purely for its early impact evidence in comparison to citation counts, emphasizing the importance of the publication lag for citations. It used reader counts for early evidence of the success of an article promotion strategy in a randomized controlled trial [29.132].

29.6.8 Illustration of Mendeley Data

Three documents were compared with Mendeley data to illustrate some of the features of Mendeley and some of the factors that should be considered when interpreting its data. Three important scientometric papers were selected for this comparison. The first is an old *Nature* article discussing citation analysis from the perspective of non-scientometricians concerned about its uses. The second is the well-known Leiden Manifesto. The third is an article about altmetrics (Table 29.1).

All three articles have high numbers of readers and Google Scholar citations. Other factors being equal, older articles should be more cited, so it seems likely that the second and third articles, from 2015, will eventually be more cited than the first one from 2002. The two newer articles already have more Mendeley readers than the first article (The counting house). This is partly because Mendeley identified readers before citations, and so newer articles take less time to catch up with older articles in terms of reader counts. It is also partly because the counting house article was published years before Mendeley was released, so its peak years of use would have preceded the existence of substantial numbers of Mendeley users.

Mendeley data includes users' professions (Fig. 29.1). Most strikingly, the altmetric article is most used by librarians. Presumably this is due to the proliferation of altmetrics in publisher websites. In contrast, professors seem to be more concerned with traditional citation-based indicators.

Mendeley data includes users' declared country of origin or work (Fig. 29.2). It seems that some countries that are taking citation analysis seriously, such as Brazil, are not concerned with altmetrics. In contrast, Canada and The Netherlands seem to be more interested in altmetrics than citation analysis, although both countries have active researchers working in both areas. The *counting house* article seems to be particularly influential in the USA, but it is not clear why.

There are substantial disciplinary differences in the uptake of the articles (Fig. 29.3). The altmetrics article has attracted the most attention in the social sciences and computer science, although both categories might be due to library and information science researchers, since this field falls within both. The citation analysis articles are of interest in the agricultural and biological sciences. This is unsurprising given the origins of the San Francisco Declaration on Research Assessment (DORA) within the life sciences (the American Society for Cell Biology), indicating an unease with misuses of citation analysis within this discipline. Figure 29.3 also confirms that all three articles have attracted substantial interest outside of their home disciplines.

29.6.9 Investigating Science with Online Reference Manager Data

An interesting application of readership data is to track the flow of knowledge between fields. There is a long tradition of using citations to track knowledge flows by interpreting a citation from an article in field A to a paper in field B as knowledge flowing from B to A [29.133]. The same is possible for readership data when the domains of the readers of an article are known.

Table 29.1 Mendeley readers and Google Scholar citations for three scientometrics articles

Title and authors	Year	Readers	GS cites	Reads/cites
Citation analysis: The counting house *by* Adam, D.	2002	173	576	0.30
The Leiden Manifesto for research metrics *by* Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., Rafols, I.	2015	634	474	1.34
Do altmetrics correlate with citations: Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective *by* Costas, R., Zahedi, Z., Wouters, P.	2015	389	251	1.55

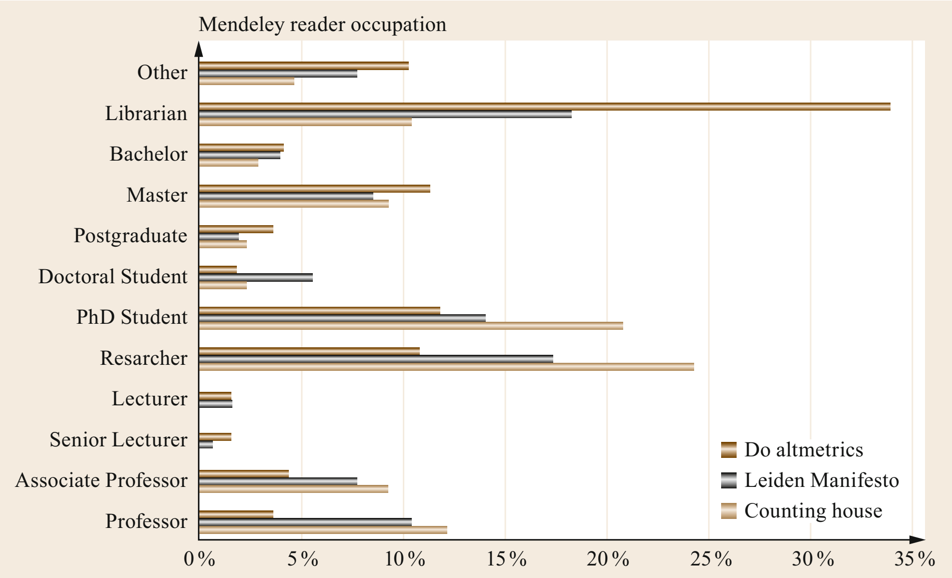


Fig. 29.1 All Mendeley reader occupations for three scientometric articles

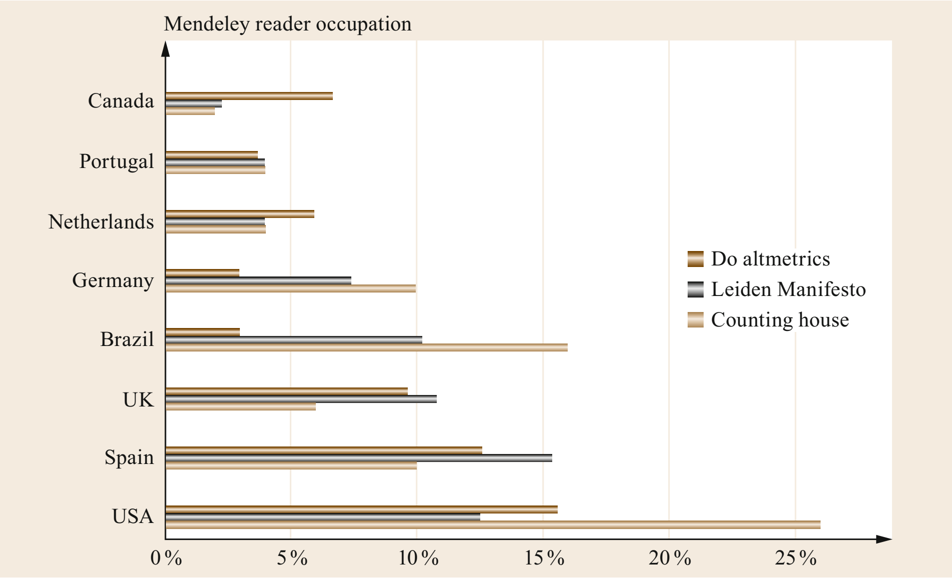


Fig. 29.2 Countries of Mendeley readers for countries with at least ten readers

The advantages of using readership data for this are timeliness and its ability to capture slightly wider impact because of the inclusion of students. A lot of data is needed to give good results however, which was a problem with one CiteULike study [29.134]. An investigation comparing knowledge flows based on Mendeley readership data with citation-based knowledge flows found differences suggesting that researchers in some fields, including business, read widely but cited narrowly [29.105].

A related application is the discovery of research clusters by identifying groups of articles read by the same user and then clustering them based on co-readership information [29.135], although this seems to be no longer possible with Mendeley.

Readership data has also been used to investigate academics through their reference lists, when these are public [29.92], to evaluate journals through the extent to which they are read [29.78] and to support literature search systems [29.136].

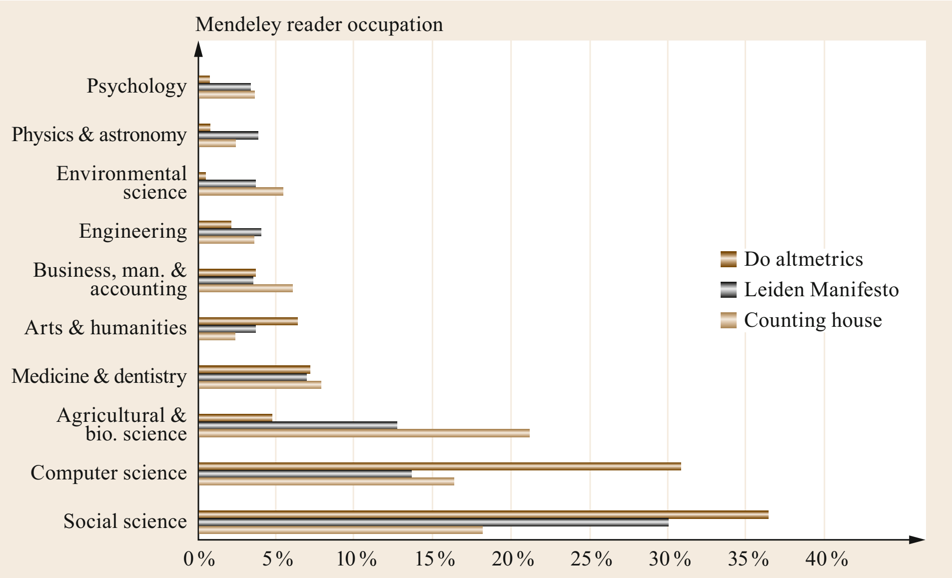


Fig. 29.3 Disciplines of Mendeley readers for the top ten disciplines

29.6.10 Advantages and Disadvantages of Reference Manager Data Compared to Citation Counts

The timeliness, wider impact, and reader demographic information advantages of readership data from all sources have already been mentioned, as have the disadvantages that it is sometimes not transparent and always open to manipulation, with a biased user base. Some additional factors are important to consider when evaluating readership data. The advantages are listed first:

- Traditional citation indexes, such as WoS and Scopus, have national biases and limits in coverage [29.137] whereas there are no restrictions on the articles that may be added to reference managers.
- Readership data is often free whereas citation indexes, except Google Scholar and Microsoft Academic, tend to charge for access.
- Readership data is relatively easy to access on a large scale from sites with an API. For example, the free Webometric Analyst software can download Mendeley records via its API from the articles' DOIs and/or metadata.
- Readership data tends to be more numerous than citation counts (e.g., [29.138]), except for older articles, and tests using it can, therefore, be statistically more powerful.

There are also additional disadvantages with readership data:

- Whereas, in theory, it is possible to find out how a work has been cited by reading the text accompanying the citation, references are rarely annotated with information that reveals why a publication was selected. Reviews are annotated readings, and these are available from sites like Goodreads and Amazon for books [29.108, 139].
- Despite the recognized national biases in citation indexes, Mendeley readership data seems to be more nationally biased than citation counts [29.140].
- Some altmetric sources of readership data can give inconsistent results [29.141], and there is a need for standardization between data providers and sources [29.142].
- Younger readers are more represented in Mendeley [29.104], and the share of younger readers may vary by narrow field and publication year.
- Differences in adoption levels and behaviors across disciplines is a complicating factor when interpreting the results of any multidisciplinary analysis [29.83].
- Some publication information entered by users to record their references is incomplete, leading to missed data [29.143]. This may be more frequent for documents with mathematical titles or in languages that are not represented by the ASCII character set.

29.7 Usage Data from Academic Social Network Sites

The online environment for science communication is continually evolving, and usage data is now not only available from publishers, academic repositories, and reference managers but also from some academic social network sites. Both Academia.edu and ResearchGate allow members to log their own papers and add them to their profile pages (as Mendeley also does now). They also provide usage data on these records in the form of download or view counts. They differ from reference managers by focusing on each author's own publications rather than their references of (presumably) mainly other scholars' works. Thus, their usage data has essentially the same nature as the download and access statistics of publishers or repositories, even though their appearance is more like Mendeley. Academic social network sites are in competition with publishers as sources of published academic research and, because of this, undermine the comprehensiveness of publisher data, apparently irrespective of copyright concerns [29.144].

For research evaluation purposes, academic social network sites are not good sources of usage indicators because they have an incomplete collection of articles and do not make their usage data easily available for researchers. Nevertheless, they are important because they have many members and their scores are apparently taken seriously by many researchers [29.145–147].

ResearchGate article views correlate positively with Scopus citations and seem to reflect a wider set of users than publishing academics, putting them on a par with other sources of usage data [29.148]. ResearchGate also provides citation counts for uploaded articles by extracting citations from all articles uploaded to the site. Although it indexes fewer citations than Google Scholar, it finds more early citations than WoS and Scopus, suggesting that many authors upload preprints to the site [29.149]. There are differing national levels of uptake of ResearchGate, which will bias its data, but despite being a type of web social network site, its data does not seem to favor younger users [29.150].

There is less research about Academia.edu, but, like ResearchGate, its scores seem to favor senior academics. They also tend to favor women, perhaps due to their greater communication expertise in the social web [29.151].

29.8 Summary

This chapter has summarized research into readership data, including usage data, with a focus on research evaluation applications but also covering collection development applications. In theory, these data are preferable to citation counts because they capture more uses of scholarly documents, such as from students and professionals. Although there is a little evidence to support his conjecture, readership data seem to primarily reflect scholarly uses in most fields. Both readership data from reference managers and usage (download/view) data from publishers have the advantage of giving early impact evidence compared to citations because of the delays associated with the publication cycle. This is due to articles being read a year or more before the citations generated by the reading, if any, appear in a citation index. Nevertheless, both download data and reference manager data can be manipulated, and whilst they are useful for informal evaluations and investigations into science itself, they should not be used for formal evaluation when those assessed can influence the data.

Readership data from reference managers have the additional promise that they can reveal something about the demographics of the readers, including their discipline, nation and job type. This can help with investigations of science communication. Download data in many cases have the practical limitation that a set of articles may originate from many different publishers, which complicates accessing them, and the data may not be fully comparable. In contrast, it is reasonable to collect reference manager readership data from a single site, such as Mendeley via its API, making them a practical source of readership information.

References

29.1 M.J. Kurtz, J. Bollen: Usage bibliometrics, Annu. Rev. Inf. Sci. Technol. **44**(1), 1–64 (2010)

29.2 S. Haustein: Readership metrics. In: *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*, ed. by B. Cronin, C. Sugimoto (MIT Press, Cambridge 2014) pp. 327–344

29.3 M. Thelwall, K. Kousha: ResearchGate: Disseminating, communicating, and measuring Scholar-

ship?, J. Assoc. Inf. Sci. Technol. **66**(5), 876–889 (2015)
29.4 C. Tenopir, D.W. King, M.T. Clarke, K. Na, X. Zhou: Journal reading patterns and preferences of pediatricians, J. Med. Libr. Assoc. **95**(1), 56–63 (2007)
29.5 F.A. McAlister, I. Graham, G.W. Karr, A. Laupacis: Evidence-based medicine and the practicing clinician, J. Gen. Int. Med. **14**(4), 236–242 (1999)
29.6 L.M. Schilling, J.F. Steiner, K. Lundahl, R.J. Anderson: Residents' patient-specific clinical questions: opportunities for evidence-based learning, Acad. Med. J. Assoc. Am. Med. Coll. **80**(1), 51–56 (2005)
29.7 A. Haque, P. Ginsparg: Positional effects on citation and readership in arXiv, J. Am. Soc. Inf. Sci. Technol. **60**(11), 2203–2218 (2009)
29.8 A. Grabowsky: Expanding access: An evaluation of ReadCube access as an ILL alternative, Med. Ref. Serv. Q. **35**(1), 16–26 (2016)
29.9 E. Emrani, A. Moradi-Salari, H.R. Jamali: Usage data, e-journal selection, and negotiations: An Iranian consortium experience, Ser. Rev. **36**(2), 86–92 (2010)
29.10 A.F.J. van Raan: In matters of quantitative studies of science the fault of theorists is offering too little and asking too much, Scientometrics **43**(1), 129–139 (1998)
29.11 J. Duy, L. Vaughan: Can electronic journal usage data replace citation data as a measure of journal use? An empirical examination, J. Acad. Librar. **32**(5), 512–517 (2006)
29.12 I. Rowlands, D. Nicholas: The missing link: Journal usage metrics, Aslib Proc. **59**(3), 222–228 (2007)
29.13 P.L.K. Gross, E.M. Gross: College libraries and chemical education, Science **66**(1713), 385–389 (1927), https://doi.org/10.1126/science.66.1713.385
29.14 H.D. White, K.W. McCain: Bibliometrics, Annu. Rev. Inf. Sci. Technol. **24**, 119–186 (1989)
29.15 L. Egghe, R. Rousseau: The influence of publication delays on the observed aging distribution of scientific literature, J. Am. Soc. Inf. Sci. **51**(2), 158–165 (2000)
29.16 D. King, C. Tenopir, M. Clarke: Measuring total reading of journal articles, D-Lib Mag. (2006), https://doi.org/10.1045/october2006-king
29.17 P.M. Davis: Information-seeking behavior of chemists: A transaction log analysis of referral URLs, J. Am. Soc. Inf. Sci. Technol. **55**(4), 326–332 (2004)
29.18 T. Stankus, B. Rice: Handle with Care, Collect. Manag. **4**(1/2), 95–110 (1982)
29.19 M.Y. Tsay: The relationship between journal use in a medical library and citation use, Bull. Med. Libr. Assoc. **86**(1), 31–39 (1998)
29.20 N.J. Butkovich: Use studies: A selective review, Libr. Resour. Tech. Serv. **40**(4), 359–368 (1996)
29.21 R.N. Kostoff: The principles and practices of peer review, Sci. Eng. Ethics **3**(1), 19–34 (1997)
29.22 P. Wouters: Citation cycles and peer review cycles, Scientometrics **38**(1), 39–55 (1997)
29.23 U.W. Jayasinghe, H.W. Marsh, N. Bond: A multilevel cross-classified modelling approach to peer review of grant proposals: The effects of assessor and researcher attributes on assessor ratings, J.R. Stat. Soc. A (Stat. Soc.) **166**(3), 279–300 (2003)
29.24 HEFCE: *Research Excellence Framework*, Vol. 2011 (HEFCE, London 2011)
29.25 Advisory Board for the Research Councils (ABRC): *Peer Review. A Report to the Advisory Board for the Research Councils from the Working Group on Peer Review* (ABRC, London 1990)
29.26 R. Smith: Peer review: A flawed process at the heart of science and journals, J. R. Soc. Med. **99**(4), 178–182 (2006)
29.27 C.J. Lee, C.R. Sugimoto, G. Zhang, B. Cronin: Bias in peer review, J. Am. Soc. Inf. Sci. Technol. **64**(1), 2–17 (2013)
29.28 I. Welch: Referee recommendations, Rev. Financ. Stud. **27**(9), 2773–2804 (2014)
29.29 A.F.J. van Raan: The Pandora's Box of citation analysis: Measuring scientific excellence – the last evil? In: *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield*, ed. by E. Garfield, B. Cronin, H.B. Atkins (ASIS Monograph Series: Information Today, Medford 2000) pp. 301–320
29.30 L. Egghe: Mathematical relations between impact factors and average number of citations, Inf. Process. Manag. **24**(5), 567–576 (1988)
29.31 R.P. Dellavalle, L.M. Schilling, M.A. Rodriguez, H.J. Van de Sompel: Bollen: Refining dermatology journal impact factors using PageRank, J. Am. Acad. Dermatol. **57**(1), 116–119 (2007)
29.32 T. Opthof: Sense and nonsense about the impact factor, Cardiovasc. Res. **33**(1), 1–7 (1997)
29.33 S.K. Boell, C.S. Wilson: Journal impact factors for evaluating scientific performance: Use of h-like indicators, Scientometrics **82**(3), 613–626 (2010)
29.34 L. Waltman, N.J. van Eck, T.N. van Leeuwen, M.S. Visser, A.F.J. van Raan: Towards a new crown indicator: Some theoretical considerations, J. Informetrics **5**(1), 37–47 (2011)
29.35 M. Thelwall: Does Mendeley provide evidence of the educational value of journal articles?, Learn. Publ. **30**(2), 107–113 (2017)
29.36 M. Thelwall: Three practical field normalised alternative indicator formulae for research evaluation, J. Informetrics **11**(1), 128–151 (2017)
29.37 J.E. Hirsch: An index to quantify an individual's scientific research output, Proc. Natl. Acad. Sci. USA **102**(46), 16569 (2005)
29.38 R. Merton: *The Sociology of Science: Theoretical and Empirical Investigations* (Univ. Chicago Press, Chicago 1973)
29.39 M.H. MacRoberts, B.R. MacRoberts: Problems of citation analysis: A critical review, J. Am. Soc. Inf. Sci. **40**(5), 342–349 (1989)
29.40 M.H. MacRoberts, B.R. MacRoberts: Problems of citation analysis, Scientometrics **36**(3), 435–444 (1996)
29.41 R.N. Kostoff: The use and misuse of citation analysis in research evaluation, Scientometrics **43**(1), 27–43 (1998)

29.42 F. Merkx, I. van der Weijden, A.-M. Oostveen, P. van den Besselaar, J. Spaapen: *Evaluation of research in context: A quick scan of an emerging field* (Rathenau Institute, Department of Science System Assessment, The Hague 2007)

29.43 C. Schloegl, W.G. Stock: Impact and relevance of LIS journals: A scientometric analysis of international and German-language LIS journals – Citation analysis versus reader survey, J. Am. Soc. Inf. Sci. Technol. **55**(13), 1155–1168 (2004)

29.44 N.L. Bennett, L.L. Casebeer, R.E. Kristofco, S.M. Strasser: Physicians' Internet information-seeking behaviors, J. Contin. Educ. Health Prof. **24**(1), 31–38 (2004)

29.45 G. Lewison: From biomedical research to health improvement, Scientometrics **54**(2), 179–192 (2002)

29.46 W.G. Stock: The inflation of impact factors of scientific journals, ChemPhysChem **10**(13), 2193–2196 (2009)

29.47 C. Tenopir, D.W. King: *Towards Electronic Journals: Realities for Scientists, Librarians, and Publishers* (Special Libraries Association, Washington 2000)

29.48 NSF: Proposal and award policies and procedures guide; Changes to the Grant Proposal Guide (GPG) Effective January 14, 2013 Virginia. Retrieved from http://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/gpg_index.jsp (2013)

29.49 H. Piwowar: Altmetrics: Value all research products, Nature **493**(7431), 159–159 (2013)

29.50 HEFCE (2011): Decisions on assessing research impact. Higher Education Funding Council for England. https://www.ref.ac.uk/2014/media/ref/content/pub/assessmentframeworkandguidanceonsubmissions/GOS%20including%20addendum.pdf

29.51 G. Abramo, T. Cicero, C.A. D'Angelo: Assessing the varying level of impact measurement accuracy as a function of the citation window length, J. Informetrics **5**(4), 659–667 (2011)

29.52 L. Drooge, P. Besselaar, G. Elsen, M. Haas: *Evaluating the Societal Relevance of Academic Research: A Guide* (Rathenau Instituut, The Hague 2010), http://depot.knaw.nl/9026/

29.53 S.P.L. de Jong, P. van Arensbergen, F. Daemen, B. van der Meulen, P. van den Besselaar: Evaluation of research in context: An approach and two cases, Res. Eval. **20**(1), 61–72 (2011)

29.54 J. Priem, D. Taraborelli, P. Groth, C. Neylon: Altmetrics: A manifesto. (2010) Retrieved from http://altmetrics.org/manifesto

29.55 L. Czerniewicz, C. Kell, M. Willmers, T. King: Changing Research Communication Practices and Open Scholarship: A Framework for Analysis. Scholarly Communication in Africa Programme. Retrieved from https://open.uct.ac.za/handle/11427/9068 (2014)

29.56 L. Van der Vaart, M. Van Berchum, R. Bruce, M. Burgess, G. Hanganu, N. Jacobs, D. Lecarpentier, S. Pinfield, P. Stokes: 'Open' as the default modus operandi for research and higher education. European Network for Co-ordination of Policies and Programmes on e-Infrastructures (2013) Available at: http://e-infranet.eu/wp-content/uploads/2013/03/e-InfraNet-Open-as-the-Default-Modus-Operandi-for-Research-and-Higher-Education.pdf.

29.57 B. Cronin, H.W. Snyder, H. Rosenbaum, A. Martinson, E. Callahan: Invoked on the Web, J. Am. Soc. Inf. Sci. **49**(14), 1319–1328 (1998)

29.58 S. Haustein: Taking a multidimensional approach toward journal evaluation. In: *Proceedings of the ISSI Conference*, Vol. 2 (ISSI Society, Durban 2011) pp. 280–291

29.59 S. Haustein: *Multidimensional Journal Evaluation Analyzing Scientific Periodicals beyond the Impact Factor* (De Gruyter, Saur, Berlin, Boston 2012)

29.60 R. Rousseau: Exponential decline in the use distribution of medical journals, J. Documentation **56**(4), 454–455 (2000)

29.61 M.J. Kurtz, G. Eichhorn, A. Accomazzi, C. Grant, M. Demleitner, S.S. Murray, N. Martimbeau, B. Elwell: The bibliometric properties of article readership information, J. Am. Soc. Inf. Sci. Technol. **56**(2), 111–128 (2005)

29.62 COUNTER: About COUNTER. Retrieved January 1, 2017, from https://www.projectcounter.org/about (2017)

29.63 O. Pesch: Implementing SUSHI and COUNTER: A primer for librarians, Ser. Librar. **69**(2), 107–125 (2015)

29.64 H.D. White, S.K. Boell, H. Yu, M. Davis, C.S. Wilson, F.T. Cole: Libcitations: A measure for comparative assessment of book publications in the humanities and social sciences, J. Am. Soc. Inf. Sci. Technol. **60**(6), 1083–1096 (2009)

29.65 J. Bollen, H. Van de Sompel: Usage impact factor: The effects of sample characteristics on usage-based impact metrics, J. Am. Soc. Inf. Sci. Technol. **59**(1), 136–149 (2008)

29.66 T.V. Perneger: Relation between online "hit counts" and subsequent citations: Prospective study of research papers in the BMJ, Br. Med. J. **329**(7465), 546–547 (2004)

29.67 T. Brody, S. Harnad, L. Carr: Earlier Web usage statistics as predictors of later citation impact, J. Am. Soc. Inf. Sci. Technol. **57**(8), 1060–1072 (2006)

29.68 H. Chu, T. Krichel, N. Blvd: Downloads vs. citations: Relationships, contributing factors and beyond. In: *Proc. 11th Int. Soc. Scientometr. Informetr. Conf.*, ed. by H.F. Moed (CINDOC, Madrid 2007) pp. 207–215

29.69 K.-K. Yan, M. Gerstein: The spread of scientific information: insights from the web usage statistics in PLoS article-level metrics, PLoS ONE **6**(5), e19917 (2011)

29.70 H.F. Moed: Statistical relationships between downloads and citations at the level of individual documents within a single journal, J. Am. Soc. Inf. Sci. Technol. **56**(10), 1088–1097 (2005)

29.71 H.F. Moed, G. Halevi: On full text download and citation distributions in scientific-scholarly jour-

nals, J. Assoc. Inf. Sci. Technol. **67**(2), 412–431 (2016)
29.72 L. Vaughan, J. Tang, R. Yang: Investigating disciplinary differences in the relationships between citations and downloads, Scientometrics **111**(3), 1533–1545 (2017)
29.73 J. Wan, P. Hua, R. Rousseau, X. Sun: The journal download immediacy index (DII): Experiences using a Chinese full-text database, Scientometrics **82**(3), 555–566 (2010)
29.74 J.L. Wulff, N.D. Nixon: Quality markers and use of electronic journals in an academic health sciences library, J. Med. Libr. Assoc. **92**(3), 315–322 (2004)
29.75 S. Darmoni, F. Roussel: Reading factor: A new bibliometric criterion for managing digital libraries, J. Med. Libr. **90**(3), 323–327 (2002)
29.76 J. Bollen, H.M.A. Van de Sompel Rodriguez: Towards usage-based impact metrics. In: *Proc. 8th ACM/IEEE-CS Joint Conf. Digit. Libr* (ACM, New York 2008) pp. 231–240
29.77 D. Duin, D. King, P. Van Den Besselaar: Identifying audiences of e-infrastructures-tools for measuring impact, PloS one **7**(12), e50943 (2012)
29.78 S. Haustein, T. Siebenlist: Applying social bookmarking data to evaluate journal usage, J. Informetrics **5**(3), 446–457 (2011)
29.79 K. Anderson, J. Sack, L. Krauss, L. O'Keefe: Publishing online-only peer-reviewed biomedical literature: Three years of citation, author perception, and usage experience, J. Electron. Publ. (2001), https://doi.org/10.3998/3336451.0006.303
29.80 C. Zimmermann: Academic rankings with RePEc, Econometrics **1**(3), 249–280 (2013)
29.81 C. Neylon, S. Wu: Article-level metrics and the evolution of scientific impact, PLoS Biol. **7**(11), 6 (2009)
29.82 K.J. Holmberg: *Altmetrics for Information Professionals: Past, Present and Future* (Chandos, Oxford 2015)
29.83 P. Wouters, R. Costas: Users, narcissism and control – tracking the impact of scholarly publications in the 21st century. Image Rochester NY, 50 pages. Retrieved from http://www.surffoundation.nl/en/publicaties/Pages/Users_narcissism_control.aspx (2012)
29.84 J. Liu, E. Adie: Five challenges in altmetrics: A toolmaker's perspective, Bull. Am. Soc. Inform. Sci. Technol. **39**, 31–34 (2013)
29.85 M. Fenner: What can article-level metrics do for you?, PLoS Biol. **11**(10), e1001687 (2013), http://dx.plos.org/10.1371/journal.pbio.1001687
29.86 J. Priem, H.A. Piwowar, B.M. Hemminger: Altmetrics in the wild: Using social media to explore scholarly impact, Arxiv Preprint arXiv:1203.4745 (2012) Retrieved from http://arxiv.org/html/1203.4745v1
29.87 R.R. Winslow, S.L. Skripsky, S.L. Kelly: Not just for citations: Assessing Zotero while reassessing research. In: *Information literacy: Research and collaboration across disciplines*, ed. by B.J. D'Angelo, S. Jamieson, B. Maid, J.R. Walker (WAC Clearinghouse and University Press of Colorado, Fort Collins 2016) pp. 287–304
29.88 G. Eysenbach: Can tweets predict citations? Metrics of social impact based on twitter and correlation with traditional metrics of scientific impact, J. Med. Internet Res. **13**(4), e123 (2011)
29.89 D. Benz, A. Hotho, R. Jäschke, B. Krause, F. Mitzlaff, C. Schmitz, G. Stumme: The social bookmark and publication management system Bibsonomy, VLDB J. – Int. J. Very Large Data Bases **19**(6), 849–875 (2010)
29.90 R. Jäschke, A. Hotho, C. Schmitz, M. Grahl, B. Krause, G. Stumme: Organizing publications and bookmarks in BibSonomy, CEUR Workshop Proc. **273**, 2–5 (2007)
29.91 J.L. Ortega: *Social Network Sites for Scientists: A Quantitative Survey* (Chandos, Cambridge 2016)
29.92 Á. Borrego, J. Fry: Measuring researchers' use of scholarly information through social bookmarking data: A case study of BibSonomy, J. Inf. Sci. **38**(3), 297–308 (2012)
29.93 S. Reher, S. Haustein: Social bookmarking in STM: Putting services to the acid test, Online Lead. Mag. Inf. Prof. **34**(6), 34–42 (2010)
29.94 T. Bogers, A. Van Den Bosch: Recommending scientific articles using citeulike. In: *Proc. 2008 ACM Conf. Recomm. Syst. RecSys 08* (ACM, Lausanne 2008) pp. 287–290
29.95 V. Henning, J. Reichelt: Mendeley-A Last.fm for research? In: *IEEE 4th Int. Conf. eSci. (eScience'08)* (IEEE, Los Alamitos 2008) pp. 327–328
29.96 W. Jeng, D. He, J. Jiang: User participation in an academic social networking service: A survey of open group users on Mendeley, J. Assoc. Inf. Sci. Technol. **66**(5), 890–904 (2015)
29.97 W. Gunn: Social signals reflect academic impact: What it means when a scholar adds a paper to Mendeley, Inf. Stand. Q. **25**(2), 33–39 (2013)
29.98 M. Thelwall, N. Maflahi: Are scholarly articles disproportionately read in their own country? An analysis of Mendeley readers, J. Assoc. Inf. Sci. Technol. **66**(6), 1124–1135 (2015)
29.99 M. Thelwall: *Web Indicators for Research Evaluation: A Practical Guide* (Morgan Claypool, San Rafael 2017)
29.100 X. Li, M. Thelwall, D. Giustini: Validating online reference managers for scholarly impact measurement, Scientometrics **91**(2), 461–471 (2012)
29.101 J. Bar-Ilan: JASIST 2001–2010, Bull. Assoc. Inf. Sci. Technol. **38**(6), 24–28 (2012)
29.102 S. Haustein, V. Larivière, M. Thelwall, D. Amyot, I. Peters: Tweets vs. Mendeley readers: How do these two social media metrics differ?, Inf. Technol. **56**(5), 207–215 (2014)
29.103 M. Thelwall, P. Wilson: Mendeley readership altmetrics for medical articles: An analysis of 45 fields, J. Assoc. Inf. Sci. Technol. **67**(8), 1962–1972 (2016)
29.104 E. Mohammadi, M. Thelwall, S. Haustein, V. Larivière: Who reads research articles? An altmetrics analysis of Mendeley user categories, J. Assoc. Inf. Sci. Technol. **66**(9), 1832–1846 (2015)

29.105 E. Mohammadi, M. Thelwall: Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows, J. Assoc. Inf. Sci. Technol. **65**(8), 1627–1638 (2014)
29.106 B. Hammarfelt: Using altmetrics for assessing research impact in the humanities, Scientometrics **101**(2), 1419–1430 (2014)
29.107 Z. Zahedi, R. Costas, P. Wouters: How well developed are altmetrics? A cross-disciplinary analysis of the presence of "alternative metrics" in scientific publications, Scientometrics **101**(2), 1491–1513 (2014)
29.108 K. Kousha, M. Thelwall: Can Amazon.com reviews help to assess the wider impacts of books?, J. Assoc. Inf. Sci. Technol. **67**(3), 566–581 (2016)
29.109 P. Sud, M. Thelwall: Evaluating altmetrics, Scientometrics **98**(2), 1131–1143 (2014)
29.110 X. Li, M. Thelwall: F1000, Mendeley and traditional bibliometric indicators. In: *17th Int. Conf. Sci. Technol. Ind*, Vol. 2 (2012) pp. 451–551
29.111 M. Thelwall: Do Mendeley reader counts indicate the value of arts and humanities research?, J. Librariansh. Inf. Sci. (2017), https://doi.org/10.1177/0961000617732381
29.112 M. Thelwall: Are Mendeley reader counts useful impact indicators in all fields?, Scientometrics **113**(3), 1721–1731 (2017)
29.113 K.J. Aduku, M. Thelwall, K. Kousha: Do Mendeley reader counts reflect the scholarly impact of conference papers? An investigation of computer science and engineering, Scientometrics **112**(1), 573–581 (2017)
29.114 HEFCE: *The Metric Tide Supplementary Report II: Correlation analysis of REF2014 scores and metrics* (HEFCE, London 2015)
29.115 M. Thelwall, N. Maflahi: Guideline references and academic citations as evidence of the clinical value of health research, J. Assoc. Inf. Sci. Technol. **67**(4), 960–966 (2016)
29.116 K. Weller, I. Peters: Citations in Web 2.0. In: *Science and the Internet*, ed. by A. Tokar, M. Beurskens, S. Keuneke, M. Mahrt, I. Peters, C. Puschmann, K. Weller (Düsseldorf Univ. Press, Düsseldorf 2012) pp. 209–222
29.117 D. MacMillan: Mendeley: Teaching scholarly communication and collaboration through social networking, Libr. Manag. **33**(8/9), 561–569 (2012)
29.118 E. Mohammadi, M. Thelwall, K. Kousha: Can Mendeley bookmarks reflect readership? A survey of user motivations, J. Assoc. Inf. Sci. Technol. **67**(5), 1198–1209 (2016)
29.119 M. Thelwall: Why do papers have many Mendeley readers but few Scopus-indexed citations and vice versa?, J. Librariansh. Inf. Sci. **49**(2), 144–151 (2017)
29.120 V. Larivière, C.R. Sugimoto, P. Bergeron: In their own image? A comparison of doctoral students' and faculty members' referencing behavior, J. Am. Soc. Inf. Sci. Technol. **64**(5), 1045–1054 (2013)
29.121 C. Tenopir, D.W. King, J. Spencer, L. Wu: Variations in article seeking and reading patterns of academics: What makes a difference?, Libr. Inf. Sci. Res. **31**(3), 139–148 (2009)
29.122 S. Korobili, A. Malliari, S. Zapounidou: Factors that influence information-seeking behavior: The case of Greek graduate students, J. Acad. Librariansh. **37**(2), 155–165 (2011)
29.123 E. Whitmire: Disciplinary differences and undergraduates' information-seeking behavior, J. Am. Soc. Inf. Sci. Technol. **53**(8), 631–638 (2002)
29.124 A. Mas-Bleda, M. Thelwall, K. Kousha, I.F. Aguillo: Do highly cited researchers successfully use the social web?, Scientometrics **101**(1), 337–356 (2014)
29.125 N. Maflahi, M. Thelwall: When are readership counts as useful as citation counts? Scopus versus Mendeley for LIS journals, J. Assoc. Inf. Sci. Technol. **67**(1), 191–199 (2016)
29.126 M. Thelwall, P. Sud: Mendeley readership counts: An investigation of temporal and disciplinary differences, J. Assoc. Inf. Sci. Technol. **67**(12), 3036–3050 (2016)
29.127 Z. Zahedi, R. Costas, P. Wouters: Do Mendeley readership counts help to filter highly cited WoS publications better than average citation impact of journals (JCS)? In: *Proc. 15th Int. Conf. Int. Soc. Scientometr. Informetr*, ed. by A.A. Salah, Y. Tonta, A.A.A. Salah, C. Sugimoto, U. Al (Boğaziçi University Printhouse, Istanbul 2015) pp. 16–25
29.128 N. Maflahi, M. Thelwall: How quickly do publications get read? The evolution of Mendeley reader counts for new articles, J. Assoc. Inf. Sci. Technol. **69**(1), 158–167 (2018)
29.129 R. Fairclough, M. Thelwall: National research impact indicators from Mendeley readers, J. Informetrics **9**(4), 845–859 (2015)
29.130 M. Thelwall, K. Kousha, A. Dinsmore, K. Dolby: Alternative metric indicators for funding scheme evaluations, Aslib J. Inf. Manag. **68**(1), 2–18 (2016)
29.131 A. Dinsmore, L. Allen, K. Dolby: Alternative perspectives on impact: The potential of ALMs and altmetrics to inform funders about research impact, PLoS Biol. **12**(11), e1002003 (2014)
29.132 P. Kudlow, M. Cockerill, D. Toccalino, D.B. Dziadyk, A. Rutledge, A. Shachak, G. Eysenbach: Online distribution channel increases article usage on Mendeley: A randomized controlled trial, Scientometrics **112**(3), 1537–1556 (2017)
29.133 H. Small: Update on science mapping: Creating large document spaces, Scientometrics **38**(2), 275–293 (1997)
29.134 J. Jiang, D. He, C. Ni: Social reference: Aggregating online usage of scientific literature in CiteULike for clustering academic resources. In: *Proc. 11th Annu. Int. ACM/IEEE Jt. Conf. Digit. Libr. – JCDL '11* (ACM, New York 2011) p. 401
29.135 P. Kraker, C. Körner, K. Jack, M. Granitzer: Harnessing user library statistics for research evaluation and knowledge domain visualization. In: *Proc. 21st Int. Conf. Companion World Wide Web WWW 12 Companion* (ACM, Lyon 2012) p. 1017

29.136 T. Heck, I. Peters, W.G. Stock: Testing collaborative filtering against co-citation analysis and bibliographic coupling for academic author recommendation. In: *Proc. 3rd ACM Workshop Recomm. Syst. Soc. Web*, ed. by B. Mobasher, J. Burke (ACM, New York 2011) pp. 16–23
29.137 É. Archambault, D. Campbell, Y. Gingras, V. Larivière: Comparing bibliometric statistics obtained from the Web of Science and Scopus, J. Assoc. Inf. Sci. Technol. **60**(7), 1320–1326 (2009)
29.138 M. Thelwall: Microsoft Academic: A multidisciplinary comparison of citation counts with Scopus and Mendeley for 29 journals, J. Informetrics **11**(4), 1201–1212 (2017)
29.139 K. Kousha, M. Thelwall, M. Abdoli: Goodreads reviews to assess the wider impacts of books, J. Assoc. Inf. Sci. Technol. **68**(8), 2004–2016 (2017)
29.140 A. Mas-Bleda, M. Thelwall: Can alternative indicators overcome language biases in citation counts? A comparison of Spanish and UK research, Scientometrics **109**(3), 2007–2030 (2016)
29.141 M. Fenner: Altmetrics and other novel measures for scientific impact. In: *Opening Science*, ed. by S. Bartling, S. Friesike (Springer, Cham 2014) pp. 179–189
29.142 S. Konkiel: Altmetrics: A 21st-century solution to determining research quality, Online Search, https://scholarworks.iu.edu/dspace/handle/2022/17147, July/August (2013)
29.143 L. Bornmann, R. Haunschild: Which people use which scientific papers? An evaluation of data from F1000 and Mendeley, J. Informetrics **9**(3), 477–487 (2015)
29.144 H.R. Jamali: Copyright compliance and infringement in ResearchGate full-text journal articles, Scientometrics **112**(1), 241–254 (2017), https://doi.org/10.1007/s11192-017-2291-4
29.145 K. Jordan: Exploring the ResearchGate score as an academic metric: Reflections and implications for practice. In: *Quantifying and Analysing Scholarly Communication on the Web (ASCW'15), Oxford* (2015)
29.146 K. Jordan: *Understanding the structure and role of academics' ego-networks on social networking sites*, PhD Thesis (The Open University, Milton Keynes 2017), http://oro.open.ac.uk/48259/
29.147 E. Orduna-Malea, A. Martín-Martín, M. Thelwall, E. Delgado López-Cózar: Do ResearchGate Scores create ghost academic reputations?, Scientometrics (2017), https://doi.org/10.1007/s11192-017-2396-9
29.148 M. Thelwall, K. Kousha: ResearchGate articles: Age, discipline, audience size, and impact, J. Assoc. Inf. Sci. Technol. **68**(2), 468–479 (2017)
29.149 M. Thelwall, K. Kousha: ResearchGate versus Google Scholar: Which finds more early citations?, Scientometrics (2017), https://doi.org/10.1007/s11192-017-2400-4
29.150 M. Thelwall, K. Kousha: Web indicators for research evaluation, part 1: Citations and links to academic articles from the web, Prof. Inf. **24**(5), 587–606 (2015)
29.151 M. Thelwall, K. Kousha: Academia.edu: Social network or academic network?, J. Assoc. Inf. Sci. Technol. **65**(4), 721–731 (2014)

Ehsan Mohammadi
School of Library and Information Science
University of South Carolina
Columbia, SC, USA
ehsan2@sc.edu

Ehsan Mohammadi is an Assistant Professor at the iSchool at the University of South Carolina, USA. Ehsan holds a PhD in Information Science from the Statistical Cybermetrics Research Group at the University of Wolverhampton, UK. He has served as a reviewer for several journals. His research interests are scholarly communication, social media, online information behavior, open science, information visualization, and health informatics.

Mike Thelwall
Faculty of Science and Engineering
University of Wolverhampton
Wolverhampton, UK
m.thelwall@wlv.ac.uk

Mike Thelwall, Professor of Data Science, leads the Statistical Cybermetrics Research Group at the University of Wolverhampton, UK, which he joined in 1989. He is also Docent at the Department of Information Studies at Åbo Akademi University and a research associate at the Oxford Internet Institute. His received his PhD was in Pure Mathematics from the University of Lancaster. Mike has developed tools and methods. He is an Associate Editor of the Journal of the Association for Information Science and Technology and sits on the Editorial Boards of the Journal of Informetrics, Scientometrics, and the Journal of Information Science. He was a member of the UK's independent review of the role of metrics in research assessment (2014–15) and is a member of the UK Forum for Responsible Research Metrics.

30. Data Collection from the Web for Informetric Purposes

Judit Bar-Ilan

This chapter reviews the development of data collection procedures on the web with an emphasis on current practices, data cleansing and matching, data quality and transparency. There are several issues to be considered when collecting data from the web. Transparency is essential to know what is included in the data source, how recent and comprehensive the data are, what timeframe is covered etc. Data quality relates to reliability and accuracy. Mistakes are inevitable, data providers, aggregators, and researchers all make mistakes, but these mistakes should be reduced to a minimum so that meaningful conclusions may be reached from the data analysis. Extensive data cleansing before starting the analysis is needed to try to correct mistakes in the data. When several data sources are used, data from different sources should be matched, and duplicates should be removed.

30.1 Background

Until the 1980s—not all that long ago—data collection was mostly done manually from printed sources. The science citation index (SCI), originally produced by the Institute for Scientific Information (ISI), first became available on CD-ROM (compact disc read-only memory) in 1988, and in 1997 the ISI Citation Indexes were rebranded as Web of Science (WoS) and became available on the web. This was the start of a new era, with not only a few institutions receiving data on magnetic tapes, but many researchers and decision-makers with subscriptions to the WoS getting easy access to a wealth of data. Other bibliographic databases also became available on the web. At the same time the Internet and the web also served as data sources, and informetric methods were applied to analyze data collected from the web. Similarities between the web-graph and citation networks were found. This led to the development of a new branch of informetrics: webometrics. Data for webometric purposes are mainly collected through either crawling or utilizing web search engines. In 2004 two new citation databases were launched, Scopus and Google Scholar, introducing new data collection challenges and opportunities (later, Microsoft also joined the bibliometric data providers by launching Microsoft Academic Search). At the same time the web has become more

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_30

and more interactive (also known as Web 2.0) allowing easy creation of user-generated content (e. g., blogs), having commenting capabilities, online reference managers and social media services, academic and not necessarily academic. These new web-based platforms are also used for disseminating and discussing academic research. The discussions, comments, tweets, likes, followers, views and downloads, postpublication online peer review, readership counts, etc. can serve as early signals of appreciation and impact as opposed to citations in peer-reviewed publications which take much longer. In addition, not all users contributing to these activities are authors of future publications, and therefore analysis based on Web 2.0 data sources supplement the traditional citation-based methods in bibliometrics and research evaluation. This youngest branch of informetrics was named altmetrics in 2010. The new data sources often have APIs (application programming interfaces) that allow downloading of large quantities of data. In addition to the raw altmetric data sources, there are also aggregators like http://Altmetric.com and PlumX.

30.2 Early Studies

The World Wide Web (web in the following) was founded in 1989 by Tim Berners-Lee and the first textual web browser was introduced in 1991 [30.1–3]. The web really started to take off around 1995, when graphical browsers were introduced and commercialization began.

One of the first studies on web data utilizing bibliometric methods was presented by *Ray Larson* [30.4]. He was interested in pages on geography, and used the then state-of-the art search engine, AltaVista to retrieve pages on the topic. One hundred fifteen results were listed. After the author reviewed the relevance of the results, about 40 pages remained. Links from these pages were extracted and filtered to include only pages judged most relevant to the topic. This procedure resulted in 34 core documents and a *colink* analysis was conducted, utilizing AltaVista's backlink search feature. Two sites, A and B are colinked if there is a third site, C, that links to both sites, and the strength of the colinkage is a function of the number of sites that link both to A and B. To create the raw colinkage matrix, links to all pairs of core documents were searched. It took 5 hours to complete the data collection process. It should be noted that the link: feature of AltaVista and later of Yahoo! was an excellent feature for bibliometric and webometric purposes, but neither of these search features exist anymore. Google never had a comprehensive backlink search, and now the link: search operator has been seemingly depreciated, since it does not appear in the list of current search operators [30.5].

The next milestone was two papers by *Peter Ingwersen*, the first one with his student, the late *Tomas Almind* [30.6], where they showed that informetric methods can be utilized on the web. They compared references in scholarly papers to links on the web, showed both the similarities and the differences between the two ("citation databases are retrospective, whereas the web is constantly in real time" [30.6, p. 406]). As a case study, they compared Denmark's share of the web to other Nordic countries. For data collection, they used now nonexistent sources: the search engines Lycos and AltaVista, and the Nordic web index. Data were collected several times over 2 years and growth trends were tracked. Other characteristics of the web pages were also shown. The second major contribution of *Peter Ingwersen* [30.7] to webometrics (roughly defined as informetric methods applied to the web) was the definition of the WIF (web impact factor) as an analogue to the journal impact factor (JIF). The definition of the WIF is the number of link pages pointing to the entity (site or country) divided by the size (number of pages) of the entity on the web.

Alistair Smith [30.8] refined the concept by differentiating between external links (links from another domain or site) and selflinks (links emanating from the given web site or domain) and claimed that external WIFs are more meaningful. He calculated the WIFs of New Zealand universities and compared them to the number of publications and to the number of publications/faculty member in the university. He found no correlation between the two. As in most of the studies at the time AltaVista was used for data collection. *Mike Thelwall* [30.9] showed that the coverage of country domains of the web by search engines is far from uniform, which makes a difference in the calculation of the WIF. His data collection strategy was querying the domain name server for domains, and then checking whether the domain is indexed by AltaVista.

Another early study [30.10] searched for names of five well-known full professors in information science, using the five major search services at the time: Excite, Infoseek, Lycos, WebCrawler, and Yahoo!. Eleven types of mentions (invocations) were defined. Slightly different search strategies were used for each search tool, to retrieve comprehensive lists of mentions. The largest category of mentions were conference-related activities.

30.3 Applying Bibliometric Laws to Data Retrieved from the Web

Ronald Rousseau [30.11] also explored sites referring to other sites (through links) and called these links "sitations", and showed that Lotka functions apply both to the domain distribution of search results and to sitations. The data collection tool was AltaVista.

At about the same time, I showed [30.12] that bibliometric laws apply to Usenet newsgroups (the forerunners of discussion lists, forum, and social media platforms). Usenet is a sort of electronic message board, with a hierarchical arrangement of the categories, called Usenet groups. A group roughly corresponds to a journal and a message to an article published in a journal. The major difference is there is no barrier to send a message to a newsgroup if the message is on topic (if the group is moderated). This is quite similar to the current discussions on social media platforms. Usenet newsgroups existed long before the web, but became easily searchable through the web using first the AltaVista search engine, and later Google. This work explored the temporal characteristics of interest in an event (the *mad cow* disease) and the distribution of Usenet messages on the topic between the different groups. The results showed that the messages were distributed more or less according to Bradford's law, and the interest in the topic rose and then died down quickly.

30.4 Longitudinal Studies

Bar-Ilan [30.13] collected information from several search engines on the late mathematician, Paul Erdős. At that time using a few tricks, it was possible to retrieve from the search engines all the results for a given query (maybe not for very popular topics). The aim of the study was to show how the great mathematician is depicted on the web. Data were collected from seven comprehensive search engines: AltaVista, Excite, Infoseek, Lycos, Magellan, Opentext, and Yahoo!. This data collection process led to the realization that the overlap between the search engines is quite small [30.14]. Today, we have only a few comprehensive search engines, and in the Western world, Google is most used, but we should keep in mind that Google does not and cannot cover everything. Studies on the overlap between search engines cannot be conducted anymore, because of the vast amount of information and because the search engines limit the number of results that they provide. In 1997 (time of data collection) altogether 6681 web pages were found containing the search term `Erdos`. As of May, 2017, Google reports that it located about 1 830 000 results, Bing reports *only* 319 000, Yandex 3 million results, and Baidu seemingly does not report the number of search results.

Bar-Ilan and *Peritz* [30.15] carried out an 8-year longitudinal study starting in 1998, using the search phrase `informetric OR informetrics`, where the query was run once a year, except for 2000 and 2001. Two data collection methods were used: search engine results and revisiting previously located web pages containing one of the search terms and located in a previous year. As pointed out before, the web is not static, new pages are created, old ones can get removed or updated (the search term does not appear anymore on the page) or moved to a new URL. In 1998, 866 pages were located, and by 2006 this number had increased 33-fold to 28 914 pages. Out of the 866 pages, only 165 pages (19%) still existed in 2006 and contained one of the search terms. The search engines used for data collection were AltaVista, Excite, HotBot, Infoseek, Lycos, and NorthernLight (all defunct now). The search engine scenery has changed considerably during the years, and in the last two years the data collection was from Google, Exalead, Teoma (Ask), and Yahoo!. Brent Payne's photostream on Flickr has nice screenshots of the user interfaces of some of these search engines (start from https://www.flickr.com/photos/brentdpayne/4306540031/in/photostream/). At the last data collection point *tricks* (query chunking) were used to overcome the limitations on the number of results displayed by the search engine, by including/excluding additional search terms [e. g., (`informetrics AND scientometrics`) and (`informetrics AND NOT scientometrics`)], limiting the query by domain or site, by filetype or by date, or a combination of the previous methods. These tactics can be useful also today when collecting large amounts of data from search engines, as was suggested also by *Thelwall* [30.16]. However, search engines, especially Google, do not like extensive data collection, and they often lock out the user from searching for a few hours and even for a whole day. A detailed methodology on how to conduct such longitudinal studies on the web appears in [30.17].

Other longitudinal studies that collected data for more than a year include the one conducted by *Koehler* [30.18] for six years that monitored a fixed set of 361 pages, the study by *Gomes* and *Silva* [30.19]—a 3-year study of the Portuguese national web, and of *Baeza-Yates* and *Poblete* [30.20] that followed the changes in the Chilean web for three years.

30.5 Search Engine Reliability and Validity

Snyder and *Rosenbaum* [30.21] identified problems with the reliability of the number of results reported by search engines. They compared the results from the then existing search engines with capability of searching for links to a specific domain. There were two such search engines at the time, HotBot and AltaVista. They showed that there are differences not only in the number of reported links, but also in the ratios, and had serious reservations about the use of data collected from web search engines for research purposes. Currently link searches are not available anymore, however the number of pages in each domain is still retrievable.

Table 30.1 shows the results retrieved on May 6, 2017 from Google and Bing for the generic top-level domains and for a few country top-level domains, English speaking and non-English speaking, questioning the validity of the reported numbers and/or the coverage of the search engines.

The longitudinal studies conducted by Bar-Ilan and Peritz, mentioned before, led to the realization that not only is the web dynamic, but so is search engines' coverage of the web. Pages indexed at one point in time might be excluded from the index a month later and reappear two months later and might disappear and reappear again. There can be several explanations for this behavior: search engines refresh their databases either periodically or continuously. If they do it periodically (which was done most probably by some of the search engines in the 1990s), they recrawl from scratch and since the storage space of a search engine is finite they may reach or not reach a given page. Such behavior was demonstrated by *Bar-Ilan* [30.22] for the search engine Excite. Other reasons for disappearance/reappearance could be that the page was temporarily not available at the time the search engine tried to visit it or other considerations by the search engine. *Bar-Ilan* [30.22] demonstrated monthly fluctuations, while in another paper [30.23], daily fluctuations were observed for 20 queries tracked for ten days. In this study two search engines were studied, HotBot and Snap, where both were powered by the Inktomi database, and in addition to the daily fluctuations in the results, there were considerable differences between the number of results reported by the search engines, that were two different interfaces to the same database. Currently, Bing powers Yahoo!'s search results and there are differences in the number of results reported by the two search engines (to be more precise, by the two search interfaces). On May 6, 2017, Yahoo! search reported about 200 000 results

Table 30.1 Sizes of generic top-level domains as reported by Google and Bing (May 6, 2017)

Top-level domain	Google	Bing
.com	25 270 000 000	27 800 000 000
.org	4 520 000 000	1 360 000 000
.edu	610 000 000	106 000 000
.net	1 300 000 000	1 880 000 000
.gov	1 280 000 000	107 000 000
.mil	83 000 000	10 400 000
.uk (United Kingdom)	1 560 000 000	123 000 000
.ca (Canada)	956 000 000	151 000 000
.au (Australia)	723 000 000	206 000 000
.nz New Zealand)	204 000 000	56 800 000
.es (Spain)	445 000 000	81 400 000
.fr (France)	1 250 000 000	154 000 000
.de (Germany)	1 690 000 000	427 000 000
.il (Israel)	113 000 000	41 700 000
.cn (China)	203 000 000	97 800 000
.ru (Russia)	442 000 000	96 900 000
.br (Brazil)	566 000 000	157 000 000
.za (South Africa)	181 000 000	3 370 000

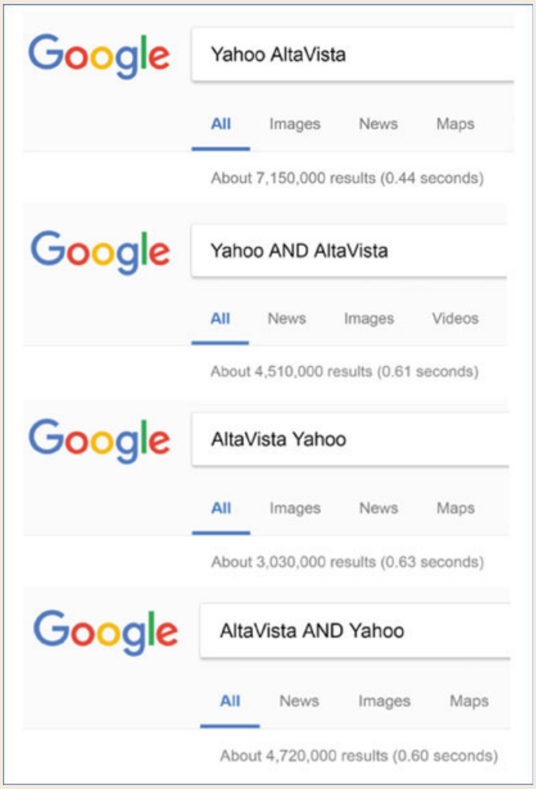


Fig. 30.1 Google search results with and without AND

and Bing reported about 429 000 results for the query `altmetrics`. Thus, one must be aware of reliability issues when using the number of results reported by search engines.

Other studies showed [30.24, 25] that for some unknown reason, search engines do not retrieve all the pages indexed by them. At that time the number of results to a query was small, thus such studies were plausible. However, the techniques applied by Bar-Ilan, pulling all search results from several search engines and checking whether the URLs not retrieved by the given search engine, but retrieved by others are in the given search engine's database, cannot be applied anymore because the number of results for an intelligent query exceeds the number of search results the search engine is willing to display, although for very specific queries the query chunking method, mentioned above, might be applicable.

Another issue regarding reliability is that search engines are not doing well in math. *Ingwersen* [30.7] noted that for the search engine AltaVista, the number of results for a query `A and B` is not identical to the number of results for a query `B and A`. Google is not better, as was shown in [30.25], and in a very recent test conducted on December 18, 2017, using http://google.com (Figs. 30.1 and 30.2). The queries were run within 10 min. The Boolean operator AND does not appear in the current help sheet of Google [30.5], it is supposed to be assumed, however when running the query with and without the AND operator, entirely different results are received as can be seen in Fig. 30.1. Among the inconsistencies is the huge difference between the results for `AltaVista AND Yahoo` versus `Yahoo AND AltaVista` (Fig. 30.1). Google also fails in the `A AND B` versus the `B AND A` test, irrespective of the use of the AND operator, and the NOT operator (- in Google). The Venn diagram in Fig. 30.3, clearly shows the impossibility of the reported number of results for the queries. Of course, Google has a good excuse, the numbers are only estimates, however these estimates do not make sense. A more plausible explanation is that these query operators are rarely used, and thus search engines do not need to make an effort to make the numbers consistent. Users rarely go beyond the first 10

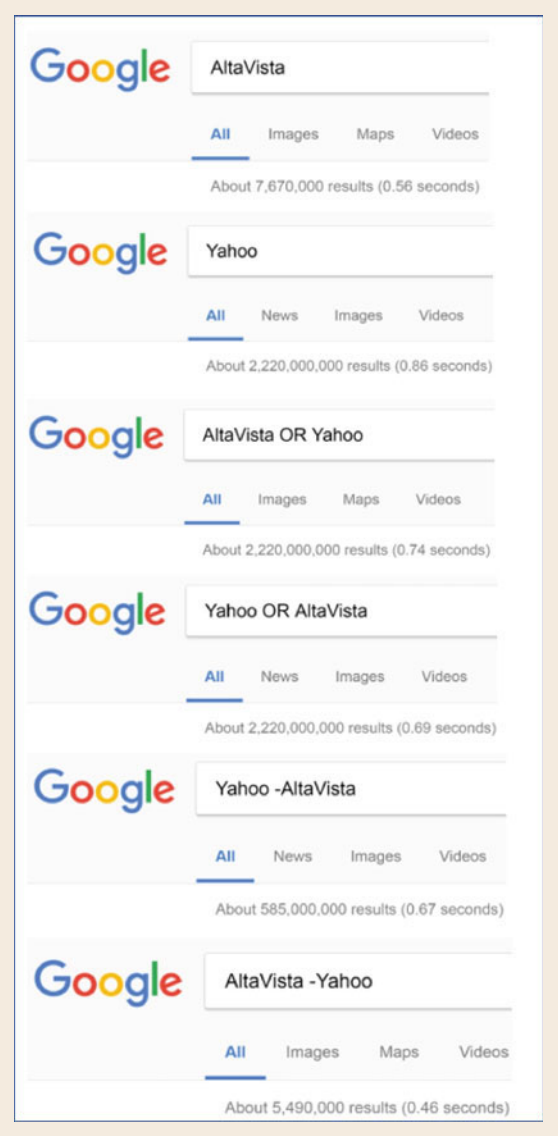


Fig. 30.2 Inconsistencies in the number of search results

or 20 search results [30.26], and do not care whether there are ten thousand or ten million results for a given query.

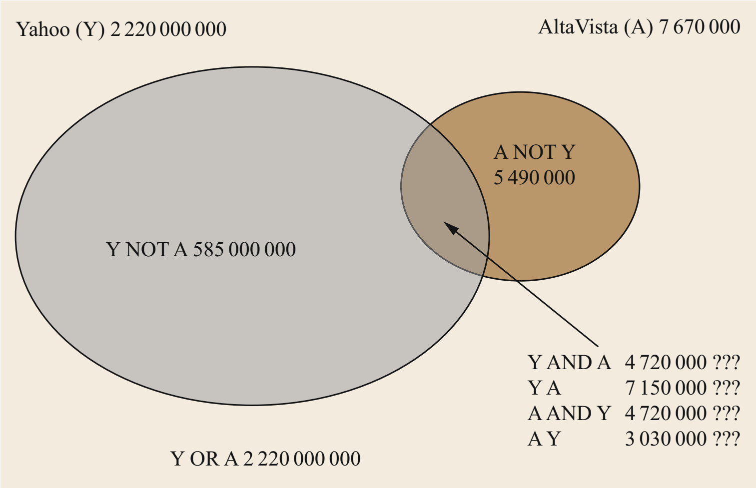


Fig. 30.3 Venn diagram of the number of search results reported for searches combining the query terms `AltaVista` and `Yahoo`

30.6 Data Cleansing

As seen above, a lot of problems in the data collection stage have been identified. These should be clearly identified in web-based studies and listed in the limitations. Steps can be taken in trying to overcome this problem by trying to cleanse the data. *Wikipedia* [30.27] defines data cleansing as

> the process of detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data.

Data cleansing is an essential step for data collection not only from the web but from other sources (like commercial databases) as well. The major difference is that the web is an uncontrolled environment, thus data cleansing is an absolute must. It is a time-consuming and iterative process, that includes deduplication (especially when data are collected from several sources), identifying outliers and deciding how to handle them, resolving inconsistencies, and correcting obvious mistakes [30.28, 29].

30.7 Link Analysis

Snyder and *Rosenbaum* [30.21] were the first to study linkage between domains, and as a test case they showed links between and within generic top-level domains. *Thelwall* [30.9] suggested that top-level domains are too coarse, and that academic domains tend to be older and better linked than commercial ones. He took his own advice and concentrated on studying the linkage between university domains mainly in the UK. In several countries academic websites have a special subdomain, like .edu in the US or .ac.uk in the UK. First, because in a previous paper [30.9] he showed that the coverage of the then existing search engines was not satisfactory, he built his own "web impact factor crawler" [30.30]. The homepage of the university website was the starting point for the crawl to cover the whole website by following links. While crawling the website, all links were collected. This was done systematically for six universities at first and was conducted twice to prove the reliability of the data. The crawler encountered several obstacles, and sometimes manual intervention was needed to overcome them, and the collected data also had to be cleansed. Thelwall, like *Smith* [30.8] could not find a relationship between the research ranking of the university and its WIF, which is understandable because many links (even external links) are not research related. The crawler is an alternative to information retrieval from search engines, but it needs computational sources, technical capabilities that are not readily available to all researchers. It should be noted, that Mike Thelwall developed several data collection tools from the web, and keeps them updated, and most importantly these tools are freely available for use by researchers. The tools have a common interface, now called *Webometric Analyst*, and can be accessed at http://lexiurl.wlv.ac.uk/. The crawler, *SocSciBot* is also freely available from http://socscibot.wlv.ac.uk/.

Thelwall extended his initial, proof-of-concept paper, and extensively studied links between UK universities and managed to disprove initial findings [30.8, 9] that there is no relationship between conventional academic measures and WIF, by modifying the definition of the WIF and filtering out links that were not research related [30.31]. The process of categorizing links was very time consuming and somewhat subjective. The method using the crawler was compared to data collection using AltaVista. The findings were that AltaVista found more links between the university websites than the specially designed crawler and there was a strong and significant correlation between the WIFs calculated based on the two methods. It was also claimed that AltaVista was reliable, in contrast with previous findings [30.21], which can easily be explained by improvements and increased coverage of AltaVista over the years.

At first only links between universities were counted, but later the inlinks were extended to other inlink sources (.edu, .ac.uk, and .uk). The external WIF was further modified by replacing the denominator: instead of the size of the university website, the number of full time equivalent (FTE) research staff was used. Using this modified WIF strong correlations were found between the WIF and the research assessment scores from the RAE (the old UK research assessment exercise, now replaced by REF—research excellence framework) [30.32]. *Thelwall* defined additional WIFs based not on links to individual web pages but to directories or to domains, called range metrics [30.33, 34], and studied the effects of geographic distances [30.35] utilizing the link crawler developed by him. Other link analysis studies, like linguistic characteristics of linking within a set of European countries [30.36] or studying the linkage between Asia-Pacific university websites [30.37] were carried out using AltaVista's search features. A large number of additional studies on link analysis were carried out by *Mike Thelwall* and his students and collaborators, as documented on his website (http://www.scit.wlv.ac.uk/~cm1993/mtpublications.html#LinkAnalysisMethods) and summarized in his books on link analysis [30.38] and on webometrics [30.39].

A different data collection method is to explore colinks instead of direct links between entities. In this process a set of target entities are identified, and pages that link to pairs of target websites are retrieved using search engines. Such a study of Canadian universities was conducted by *Liwen Vaughan* [30.40]. In this study, Yahoo! search was used to retrieve colink data. Yahoo bought both AltaVista and AlltheWeb in 2003, and used to have extensive capabilities for retrieving inlinks and colinks.

Unfortunately, there are no commercial search engines anymore that allow retrieving links to websites, and thus alternative, but less comprehensive techniques were developed to assess the web visibility of web sites or web pages, e. g., URL citations (where the URL of the linked web page explicitly appears on the linking page) [30.41] or linked title mentions [30.42]. The link title mentions method first uses a search engine to retrieve pages that mention the name of the target website, and then using the previously mentioned *Webometric Analyst* automatically checks whether there is a link to the target website on the page.

Clearly, counting links from/to websites is not sufficient. One must try to understand the reasons for linking. *Kim* [30.43] interviewed researchers on why they inserted hyperlinks into their electronic publications, and identified three major motivations: scholarly, social, and technological. Although interviews are highly useful, they are also time consuming. An alternative technique is trying to deduce from the context of the link the reason for creating it. *Wilkinson* et al. [30.44], using a sample of about 400 pages from UK universities linking to other universities, created a typology of reasons to link, including teaching materials, information for students, research partners, research support, research reference, recreational, e-journals, and similar department. *Bar-Ilan* [30.45] claimed that, to attempt to understand the motivation for linking in the academic context, one must consider the properties of the source and the target pages, the properties and context of the link, including the relationship between the source and target page. The entries in the codebook were derived from the literature and from a small sample of linkage data between Israeli universities, and then successfully applied to a larger sample of nearly 600 pages.

Besides analyzing links to and between academic institutions, the linking structure of other types of entities was also explored. Vaughan et al. found relationships between inlink counts to commercial sites and their business performance measures. In 2004, she [30.46] studied American and Canadian IT companies. The data collection steps for such a study are clearly described, starting with selecting companies for which financial information is available and that have a presence on the web. Web data were collected from multiple search engines to minimize possible bias. Significant correlations were found between inlink counts and business performance indicators. Later colink data was used to explore similarities between companies [30.47]. Commercial companies in the same field are competitors and have no incentive to link to each other, thus third-party sites are needed to map the positions of different companies in the field. This study, like [30.40], used Yahoo! search for data collection. In order to understand the reasons for linking to commercial websites, a classification scheme was designed, and a content and context analysis of the linking websites was carried out [30.48]. As for academic inlink studies [30.44, 45], presenting numbers and relationships is not sufficient, and quantitative studies should be supplemented by qualitative ones.

Leydesdorff and *Curran* [30.49] and *Stuart* and *Thelwall* [30.50] explored university–government–industry links. In the first study the data source was AltaVista and it concentrated on Brazil and the Netherlands. In the second paper, the UK West Midland automobile industry was chosen. In this study, Stuart and Thelwall utilized the Google search API (not existent anymore). Although Mike Thelwall had his own crawler, it was not used because commercial sites often let in the crawlers of the well-known search engines, but not academic crawlers. By 2006, Google could extract links not only from html files but also from Word or PDF files. Stuart and Thelwall used query modifiers that were available at the time of the study but are not available anymore, and instead of actual links (Google never had good means for reporting inlinks to a site), the URL citation method [30.41] was applied. They were able to create meaningful visualizations of the connections between the websites from the data collected. They concluded that URL citations only partially match real-world relationships.

30.8 Bibliometric Citations Versus Web References

In an early paper, *Bar-Ilan* [30.23] showed that the web is a rich source of metadata on scholarly publications on the web. While searching for the appearance of the term `informetrics`, she found that about 40% pages included references to scholarly publications. The number of times authors, journals, and publications appeared in the retrieved web pages was calculated. This can be seen as a forerunner of altmetric counts on social media platforms, indicating visibility of publications. The coverage of the references (metadata) found on the web was compared to several commercial databases. There was a 75% overlap between WoS articles published after 1995 with the term informetrics in the title of the journal article and the metadata retrieved from the web. *Vaughan* and *Shaw* [30.51] further explored this issue on a much larger data set. They quoted Tom Wilson's post in 2002 on the SIGMETRICS discussion list [30.51, p. 1313]:

> I looked at a couple of my own papers and counted the SSCI citations and then searched for mentions of the papers on the Web—the results left me wondering whether the reliance on citation indexing as a measure of performance is now past its sell by date.

Eugene Garfield replied to this message:

> I suppose that some day the people who run google and other search engines will figure out a way to separate true research citations from mere mentions of names, but in the meantime it is really not defensible to compare information retrieval via WoS or STN or Dialog of ePsyche or whatever, to searches using google or other search engines over the Internet. That is why I don't waste my time trying to do so.

Vaughan and Shaw were challenged by Garfield's reply, and retrieved articles published in 1992 and 1997 in library and information science (LIS) journals indexed by WoS, recorded the number of citations on WoS and the number of mentions on the web by 2002. Web data were retrieved using Google (Google Scholar did not exist at the time). The average correlations between WoS citations and web citations was around 0.6 for the 1997 data set, the correlations were significant for more than 60% of the journals. In a later study, they explored web mentions (called web citations) in biology, genetics, medicine, and multidisciplinary sciences [30.52] and found significant correlations between web and WoS citations. Google search was used to retrieve data, where the query was the title of the article as a phrase search and the results retrieved were manually checked and classified.

30.9 Google Scholar

The year 2004 was another milestone in data collection from the web for informetric purposes. Two new citation databases, Google Scholar and Scopus were both launched, ending the hegemony of the Web of Science. Google Scholar was launched in beta, gathering information on scientific publications from the web, from libraries and publishers, and provided citation counts and links to the documents, some behind a paywall and others freely accessible. Unlike the commercial citation databases, Google Scholar was (and is) free to use by everyone. It also uses Google's technology to suggest corrections of typos and spelling mistakes.

The initial response of the scientific community was quite mixed. *Peter Jacsó*, in a number of papers, emphasized the shortcomings of Google Scholar [30.53, 54]. In an early paper, he mainly criticized the coverage, in the paper from 2008 he took a closer look and found inaccuracies in author names, and publication dates. On the positive side, he mentioned the improved coverage of journals, the inclusion of books, impressive geographic and language coverage and (partial) coverage of digital repositories. Others criticized Google Scholar's lack of transparency about its coverage both in terms of the list of journals covered and the document types included, a definition or explanation of what are the characteristics of scholarly documents to be included in the index, lack of information on the update frequency and the ranking of search results (at the beginning most of the time the results were ordered in descending order of citations) [30.55–57]. A good summary of Google Scholar's shortcomings appears in [30.58].

One of the common methods for evaluating a bibliographic or bibliometric database is by comparing it to other similar databases in terms of coverage, freshness, accuracy, and reliability. The obvious choice is to compare Google Scholar to the Web of Science and Scopus. *Jacsó* [30.59] did this, and clearly favored the Web of Science. *Bauer* and *Bakkalbasi* [30.60] also compared the three citation databases on JASIST articles for the years 1985 and 2000. They found that Web of Science and Scopus reported similar citation counts for the articles published in 2000, and Google Scholar reported significantly higher citation counts. This result increased interest in Google Scholar and further comparison studied were conducted.

Besides citation counts, coverage is a major issue, and can also be studied by comparing Google Scholar to other databases. *Neuhaus* et al. [30.61] compared Google Scholar to 47 other databases/publishers by drawing 50 random article titles from each database and testing whether they were indexed by Google Scholar. The lowest coverage was in humanities (10%), followed by the social sciences (39%). *Neuhaus* et al. [30.61] tested article coverage, whereas *Oppenheim* and *Norris* [30.62] were interested both in journal title coverage and article coverage of WoS, Scopus, Google Scholar, and CSA Illumina. A random sample of 380 journal titles was drawn from the list of journal titles submitted to the UK 2001 research assessment exercise to 13 subject areas (units of assessment) covering the social sciences. A further random sample of 306 articles appearing in the sampled journals were tested. Their conclusion was that Scopus had best coverage. *Schultz* [30.58] conducted ten searches (topic, journal, and author searches) in parallel on PubMed and Google Scholar, and found that Google Scholar returned more results, including grey literature, but the overlap with PubMed was small. The lack of controlled vocabulary in Google Scholar was listed as a serious obstacle for complex searches. The coverage of Compendex was compared to that of Google Scholar in engineering with a decade by decade comparison, starting with the 1950s [30.63]. The authors found that Google Scholar's coverage increased by publication year to around 90% from the 1990s onwards. The data collection method for this study was to run eight topic searches and to select 20 random articles from each decade for each of the searches.

Kousha and *Thelwall* [30.64] compared web/URL citations (described above) with WoS and Google Scholar citations on a sample of Open Access articles in several science and social science fields. Significant and strong correlations were found between WoS citations and web/URL citations. Google Scholar citation counts were higher than WoS citation counts in the social sciences and in computer science, probably because for these disciplines journals are not necessarily the primary publication venue. Books and book chapters are important for the social sciences and proceedings for computer science and these are not covered (or covered only partially) by WoS.

Another major study [30.55] considered the publications of 25 LIS faculty members and explored the citations to their work on WoS, Scopus, and Google Scholar, examining more than 10 000 citing documents and checking whether these actually cited the target article. They found that Google Scholar excelled in its coverage of proceedings and non-English journals. Their recommendation was to use all three citation databases if the aim is a comprehensive assessment of scholarly impact. Their main conclusion was that even though Google Scholar had much better coverage of proceedings papers, it does not influence the rankings of the scholars, and the effort to add data from Google

Scholar is very high. Hence, WoS together with Scopus are sufficient for ranking LIS researchers.

They complained about the amount of work they invested in the study: 100 hours on WoS, 200 on Scopus, and 3000 (!) on Google Scholar. Data from WoS and Scopus were retrieved manually and data from Google Scholar was *harvested*. Data from all three sources were checked and cleansed—I assume this was the most time-consuming part, but even so 3000 hours seems a lot. Data were collected in March 2006, and the manuscript was submitted in mid-October 2006, which means 180 days between data collection and submission or 16.7 hours daily spent on data collection and cleansing of results retrieved by Google Scholar.

Google Scholar is quite unfriendly for data collection. Luckily the problem was partially solved by Anne-Wil Harzing, who released the first version of *Publish or Perish* in 2007 [30.65]. This software organizes and reranks retrieved items according to the number of citations, and provides metrics, including total cites, h-index, and g-index. In the most recent version *Publish or Perish* also covers Google Scholar Citation Profiles and yearly citation counts of the profiled author. The output can be saved in various formats, most notably for further analysis as comma-separated values (CSV) or Excel, but different citation styles are supported as well. See [30.66] for a 10-year history of Publish and Perish. This is an indispensable tool for retrieval and basic analysis of data from Google Scholar.

This tool was used in 2008 to study author and journal impact as reflected on Google Scholar [30.57] for 20 top journals in management. Strong and significant correlations were found between the JIF and the three measures derived from Google Scholar by *Publish or Perish*: h-index, g-index, and citations per paper. A follow-up paper [30.67] further advocated the use of Google Scholar's h-index for journals over the JIF. They provided supporting evidence from 838 journals in economics and business.

In contrast, *Bornmann* et al. [30.68] compared citation counts reported by WoS, Scopus, Google Scholar, and Chemical Abstracts for 1837 articles submitted to Angewandte Chemie International Edition (accepted or rejected and published elsewhere later). They found the three fee-based services provided very similar results, however large differences were observed between the fee-based sources and Google Scholar, indicating the lack of convergent validity of Google Scholar.

Database coverage is a major factor both for the number of publications and for citations, because the database is aware only of citations from publications that are indexed by the database. The h-index [30.69] is a function of both the number of publications and the number of citations—thus the h-index calculation depends on the data source, as was shown by *Bar-Ilan* [30.70] for 40 highly cited Israeli scientists listed in the HCR (highly cited researchers) database (2001 version—http://hcr.stateofinnovation.com/page/archives) and three Israeli Nobel prize winners. The data, especially from Google Scholar, were cleansed extensively, but such a process was needed for Scopus and WoS as well. The h-indices based on Scopus and WoS were similar. In a few cases the Google Scholar-based h-index was lower than the other two h-indices, but for computer science and mathematics, for almost all scientists the Google Scholar based h-index was more than 70% higher. *Meho* and *Rogers* [30.71] computed the h-index of 22 highly cited researchers, based on WoS, Scopus, and Google Scholar. The h-index based on Google Scholar was always higher than the h-index based on the other two sources (by 59% on average).

Google Scholar has since matured, as can be seen from a study by *Anne-Wil Harzing* [30.72] of 20 Nobel prize winners in chemistry, economics, medicine, and physics. The results were stable over time, and Google Scholar had good coverage of the Nobel prize winners' publications. A follow-up study [30.73] supported these findings. Longitudinal studies are of importance as they provide information on the growth and stability of the database.

In a recent large-scale study [30.74], Google Scholar was compared to Scopus, in terms of coverage, number of citations, and indexing speed. Two journals from each of six disciplines were chosen. One hundred top-cited articles published in the 12 journals between 2010–2014 were selected. Data on the 1200 selected articles were collected in 2015. Correlations of citation counts were high and significant. Except for one journal in chemistry, citation counts on Google Scholar were higher than on Scopus. Duplication due to multiple versions was observed in less than 2% of the cases.

Google Scholar, like the other citation databases, not only reports the number of citations, but also links to the citing documents. Since there were complaints about nonexisting, "phantom" citations [30.75], *Bar-Ilan* [30.76] examined all the citing documents reported by Google Scholar to the book *Introduction to Informetrics* by Leo Egghe and Ronald Rousseau. Rather surprisingly, in less than 2% of the cases the same citing URL appeared twice (double counting) and less than 7% were content duplicates (the same citing source at different URLs). After duplicate removal, all citing documents were checked for the existence of a reference to *Introduction of Informetrics*. More than 85% of the citations were genuine. This was only a case study examining 358 citing documents, but the findings on the accuracy of Google Scholar did not support complaints by *Jacsó* [30.75] or by

Meho and *Yang* [30.55]. One of the possible reasons is that Google Scholar had improved considerably over time.

Despite the cold welcome Google Scholar received at first, it has greatly improved over time. However, this does not mean that it cannot be manipulated, as was shown by *Delgado López-Cózar* et al. [30.77]. They created and uploaded six documents by a fake author, Marco Pantani-Contador, that were linked from an institutional web page. All six documents referenced all the publications of the research group. The documents were picked up by Google Scholar and resulted in an increase of 649 citations for the three authors and an average increase of 2.3 in their h-indices. Four of the *publications* of the nonexistent Pantani-Contador are still indexed by Google Scholar (GS) as of May 18, 2017, and are picked up from the supplementary material to the article describing the experiment, as legitimate publications (Fig. 30.4), even though the paper reports that after the announcement of the results of the experiment [30.77, p. 451]:

> Google erased all traces from our false fictitious researcher Pantani-Contador as well as the GS citations profiles of the authors of this article, which were kept in quarantine for some weeks without notifying the authors at any time and then cleaned and made publicly available.

A quick test of one of the papers in the reference list of the papers *published* by the nonexistent Pantani-Contador, "Strategic knowledge maps of the techno-scientific network" [30.78], was cited by 14 publications, including three of the four *publications* of Pantani-Contador appearing in Fig. 30.4 (Fig. 30.5). Another article, "State of the library and information science blogosphere after social networks boom: A metric approach" referenced by Pantani-Contador, was cited 32 times including five citations from the fake publications (four of them from the four Pantani-Contador *publications* still indexed by Google Scholar, and the fifth from the supplementary materials to [30.77] which contains the six fake publications). Thus, the research group lead by Emilio Delgado López-Cózar still profits from the fake experiment that took place more than four years ago.

It should be noted that as of December 2017 the fulltext of the fake articles is no longer available from Google Scholar, but they still count as citations to the above-mentioned articles. The fulltext of the fake articles is accessible from https://zenodo.org/record/1043395#.WjjCbmdG2Uk.

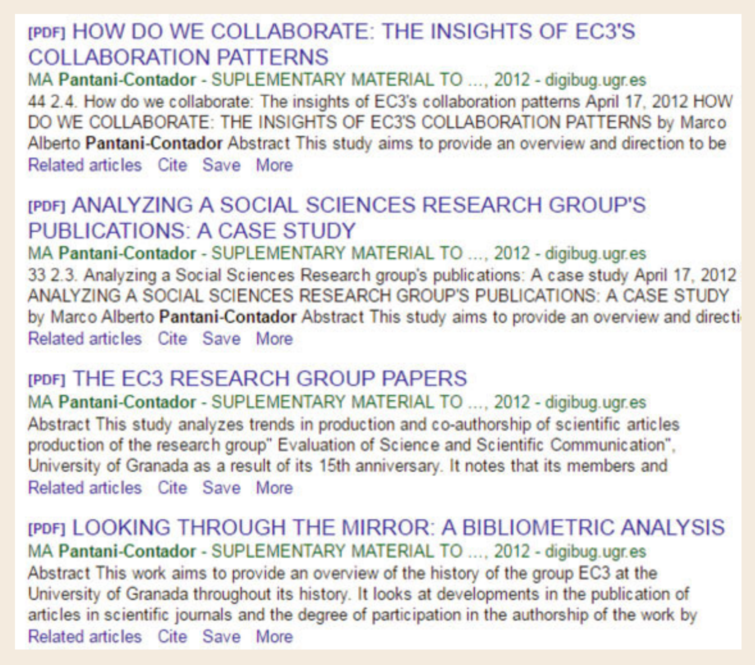


Fig. 30.4 Google Scholar continues to index fake publications, including the four above

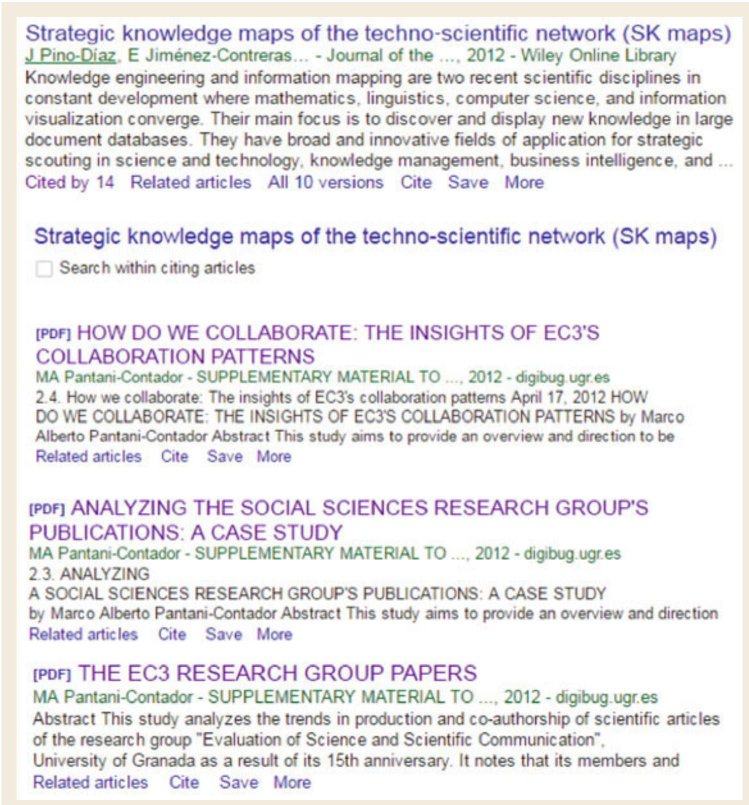


Fig. 30.5 Fake publications that continue to count as legitimate citations in Google Scholar

30.10 Additional Google Sources

An additional data source, Google Books, first named Google Print, was launched in 2002 and renamed Google Books in 2005 [30.79]. Its primary aim is to search the full text of books and magazines. Google digitizes and OCRs books by partnering with publishers and major libraries that allow Google to do that. Although Web of Science has its Book Citation Index and Scopus also indexes some books, their coverage is very poor, compared to Google Books. Google Scholar provides citation counts to books, but probably not from books. Citations to books can also be extracted from the Web of Science (using the *cited reference search*) and from Scopus (using the *view secondary documents link*), however these are not straightforward processes.

In Google Scholar, one can simply search for the book title, or for a topic, and in the search results relevant books with citation counts will appear [30.76]. The opposite direction, extracting references in books to scholarly publications has also been explored by *Kayvan Kousha* and *Mike Thelwall* [30.80], where they manually entered the titles of more than 1900 journal articles into the Google Books search box. The search results contain books that cite the given article, but also retrieve nonsense results. At first, I was very glad to see that the book search for `Which h-index`, published in 2008 retrieved about 428 results. On the first and second result pages most of the results seemed relevant, but by the fifth page my suspicions grew, as books from the 1960s and 1970s were retrieved, even though the concept of the h-index was introduced only in 2005 [30.69] and the article published in 2006. Actually, it would have been an honor to be cited by Eugene Garfield or by Derek De Solla Price. On the seventh page after 62 results, I was informed that no books contain the phrase (Fig. 30.6). The search was carried out on 19 December 2017. To overcome this problem, *Kousha* and *Thelwall* [30.81] developed an automatic method to retrieve and clean citations from books to books, a method that can be applied to searching for citations to articles from books as well and is part of the *Webometric Analyst* tool set, mentioned before.

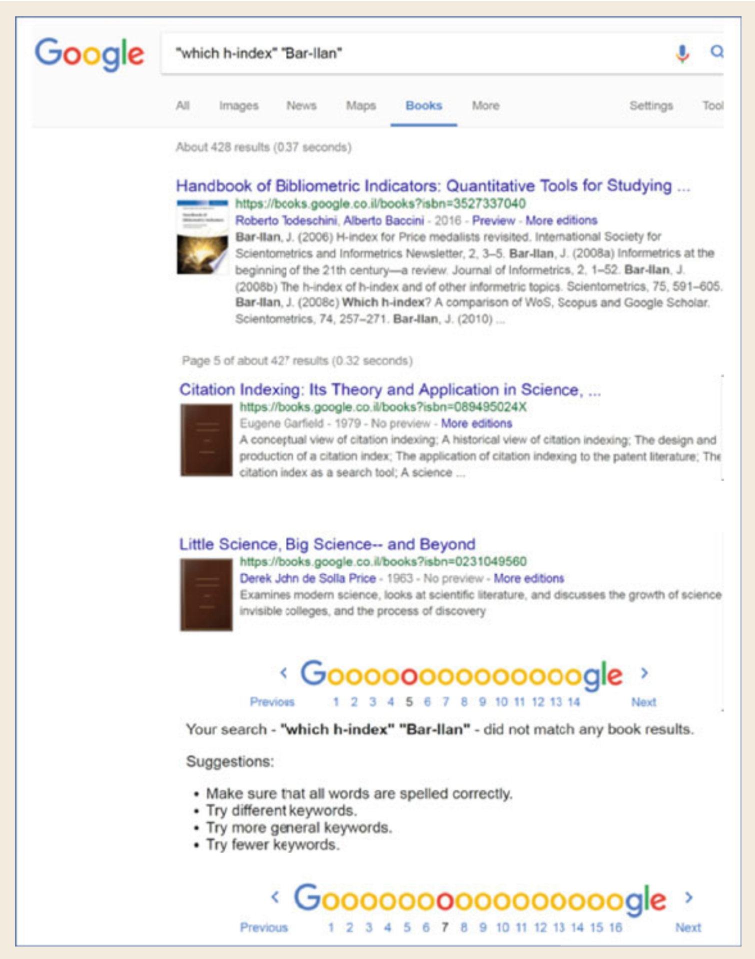


Fig. 30.6 Problems with the accuracy of Google Books citation matches

Google Scholar Citation Profiles were introduced in 2011. This tool allows researchers to set up a profile and to showcase their scholarly publications. Publications of the researcher are pulled from Google Scholar and citation counts are continuously updated. The researcher can correct the profile, delete publications that were not authored by her, and add missing ones. This tool is very popular and many institutions require faculty to setup such a profile. When setting up the profile, the researchers are asked about their affiliation. Google Scholar created institutional profiles from these data, and the Cybermetrics Lab, led by Isidro Aguillo, produces a ranking of universities based on the top Citation Profiles of researchers from the institutions. This ranking is called *Transparent Ranking* (http://www.webometrics.info/en/node/169).

Currently Google Scholar provides tools that can be utilized for research assessment. We already described Google Scholar Citation Profiles, which can be utilized in individual assessments and can be used for ranking universities. Another tool, Google Citation Metrics, launched in 2012, ranks journals according to their 5-year h-index (h-index of publications in the last five years). Each comprehensive citation database provides slightly different measures for journals: WoS's main journal measure is the Journal Impact Factor and Scopus now promotes its new measure CiteScore, however there are additional measures, like the 5-year impact factor, the eigenfactor and the article influence measure provided by WoS, and the SJR (SCImago Journal Rank) and SNIP (Source Normalized Impact per Paper) of Scopus. Scopus and Google Scholar offer their journal indicators freely, while the Journal Citation Report is accessible only to subscribers. In spite of the heavy criticism against using the impact factor for assessing individual articles (DORA Declaration [30.82], Leiden Manifesto [30.83]) it continues to have a central position in research evaluations.

The h5-index of Google Scholar has a different flavor, as can be seen by comparing the top five journals

Table 30.2 Rankings by different journal indicators

Journal	Rank—GS	Rank—IF	Rank—CiteScore
Nature	1	9	60
New England Journal of Medicine	2	2	91
Science	3	16	80
Lancet	4	4	243
Cell	5	27	17
CA—A Cancer Journal for Clinicians	NA	1	1
Nature Reviews Drug Discovery	99	3	168
Nature Biotechnology	36	5	97
Chemical Reviews	13	12	2
Annual Review of Immunology	NA	14	3
Chemical Society Reviews	6	18	5
Annual Review of Astronomy and Astrophysics	NA	10	5

ranked by impact factor (IF), h5-index, and CiteScore. Table 30.2 displays rankings published for 2015, and the differences between the most-used ranking in each database speak for themselves. Google Scholar ranks only the top 100 journals, it covers The *Annual Review of Immunology* and *CA—A Cancer Journal for Clinicians*, but their h5-indexes are lower than those in the top 100, while the *Annual Review of Astronomy and Astrophysics* is not listed at all. So far little research has been done on Google Scholar Metrics [30.77, 84, 85].

30.11 Microsoft Academic

A competitor of Google Scholar is Microsoft Academic Search. It was launched in 2006, but its development ceased and it stopped being updated in 2012. It was redesigned and relaunched in 2016 as Microsoft Academic. It is not mature enough yet, but has potential. Since it was relaunched just recently, its accuracy, coverage, and data retrieval capabilities have not been explored. This an area for future research. In July 2017 version 2.0 was launched, which seems to be a big improvement. *Harzing* and *Alakagnas* [30.86] are optimistic about the future of the relaunched Microsoft Academic. *Publish or Perish* provides an interface to download metadata from Microsoft Academic.

30.12 Subject Specific and Institutional Repositories

Subject-specific repositories like arXiv, SSRN, RePec, and PubMed Central are also a rich source for data collection and analysis, especially if they provide methods or APIs for large-scale data collection. These subject-specific repositories contain preprints (submitted for peer review), postprints (as accepted by the journal, after peer review and revision, but typically not in the journal format), and also the paper in its final format (depending on publisher permission). Postprints are especially useful for journals with long publication delays or items published in proceedings that are not online. In such cases, both the final published item and the version in the repository can accumulate citations. Early citations are directed to the repository version, but later citations may target the version on the publisher's site or the repository. It should be noted that publishers are aware of publication delays and often upload articles not yet assigned to a journal issue in an area called *early view*, but these articles are typically behind a paywall, whereas articles uploaded to repositories as pre- or postprints (depending on what the publisher allows), are green open access items with full text availability. arXiv, RePEc, and PubMed Central have APIs. *Webometric Analyst* has an interface that allows the downloading of metadata from arXiv. *Davis* and *Fromerth* [30.87] showed that depositing in arXiv gives citation advantage to the deposited articles in mathematics, while *Henk Moed* [30.88] analyzed the condensed matter section of arXiv and reached the conclusion that "ArXiv accelerates citations due to the fact that ArXiv makes papers available earlier" [30.88, p. 2047]. *Larivière* et al. [30.89] conducted an extensive comparison of arXiv and WoS, and found that 64% of the submissions on arXiv could be matched to publications in WoS. Citations from WoS to arXiv preprints were also explored. *Li* et al. [30.90] studied citations from documents indexed in Scopus to arXiv, SSRN, RePec, and PubMed Central and concluded that if a published version exists, it serves as the preferred citation target.

30.13 Altmetrics

Altmetrics are the newest addition to the metrics soup. The term was introduced in 2010 in the Altmetrics Manifesto [30.91]. The basic idea was to extend the notion of scientific impact by capturing data from the social media platforms and counting the number of relevant events [30.92]. Altmetrics events accumulate much faster than citations, and so can serve as early signals of impact. Of course, this is not always the case, because a catchy title or a controversial topic might result in a lot of likes, shares, comments, discussion, without any long-lasting impact.

Currently, the most heavily studied altmetric platforms are Mendeley and Twitter, probably because of their coverage. Mendeley and Twitter have APIs; data can be collected either directly from the APIs or by using *Webometric Analyst*. There are open source applications both in R and in Python for retrieving data using the APIs. Another option for Twitter is to request data from Altmetric.com, currently the major aggregator collecting altmetric signals of scholarly output.

Mendeley is a free online reference manager launched in 2007. When a user saves an item to her Mendeley library, she is counted as a *reader* of the item. Mendeley aggregates these reader counts and displays the number of readers of the scholarly publication. Thus, readership counts can be compared to citations. Most studies report Spearman correlations of about 0.5 between Mendeley readership counts and citation counts—showing that they are related but also different [30.93–98]. Mendeley users include authors who cite scholarly publications, but also read things that they do not cite. In addition, many users are students or other persons interested in science who have no intention to publish in scholarly venues. One should also take into account that not everything that is downloaded is read.

Twitter users are even more diverse including the general public, journal editors, and publishers who use Twitter in order to increase the visibility of their publications. *Thelwall* et al. [30.99] received data from Altmetric.com on publications that have PubMed IDs, and have at least one nonzero metric from the 11 sources covered at the time of the study by the data provider. Altmetric.com currently tracks 17 types of altmetric sources [30.100]. The aim of the study was to test whether articles with higher altmetric counts are cited more than articles with lower or nonexisting altmetric counts per altmetric. Statistically significant relationships were found between WoS citations and a number of altmetric scores, including Twitter. Because of the low coverage of the other altmetrics, only this relationship was classified as meaningful. It should be noted that as of mid-2016, Altmetric.com also tracks citations from Scopus, which should ease the process of matching citations and altmetric counts. Another more recent development is that Elsevier bought PlumX, a competitor of Altmetric, and thus it is not clear whether the collaboration between Altmetric and Scopus (owned by Elsevier) will continue. Currently, Scopus displays Mendeley readership counts for each item and additional altmetrics tracked by Altmetric.com.

Another large-scale study on tweets and citations in the biomedical literature [30.101] showed that Twitter coverage was low, and the correlation between citation counts from WoS and Twitter counts was also low. *Costas* et al. [30.102] explored the coverage of different altmetrics (Mendeley excluded) on a large data set derived from WoS. Findings show that the coverage of Twitter was highest (13%). Correlations with citation counts were low (below 0.25 for all citation measures tested), even for Twitter. For a comparison between tweets and Mendeley readers see [30.95].

It should be noted that different altmetrics reflect different aspects of *impact*. Mendeley reader counts and Reddit reflect usage, tweets probably show general interest, and other social media platforms (e.g., postpublication peer-review sites, blogs, Facebook, and academic social media sites, like ResearchGate and Academia.edu) allow for commenting and discussion. Studies were and are being conducted to explore the value of the different altmetric sources (e.g., scientific blogs [30.103]; F1000Prime [30.104, 105]; ResearchGate [30.106, 107]; Academia.edu [30.108]). As a result of an initiative of Springer-Nature and Altmetric, Bookmetrix was developed which tracks "citations, online mentions, book reviews and downloads" [30.109] both at the book- and the book-chapter level. It currently covers Springer books only, but hopefully wider coverage will be provided in the future. The metrics for each book are displayed, but access to the Bookmetrix platform is currently limited to Springer staff only. The metrics of Bookmetrix are especially useful for evaluations in the social sciences and the humanities, where in many areas traditional metrics have little value.

The Altmetrics Manifesto [30.91] also calls for measuring scholarly outputs besides articles, for example data reuse and the use of software tools. Infrastructure for measuring data reuse is being developed by DataCite, primarily by assigning DOIs (digital object identifiers) to data sets and developing a metadata schema for accurate identification of resources. Data sharing and reuse are not the norm yet [30.110], but they are encouraged by open science initiatives [30.111].

An increasing number of publishers release usage data. PLoS (Public Library of Science) was probably the first to publish article-level metrics (ALM)—see [30.112]. The idea is to be diverse and multifaceted. PLoS ALM have three central goals: to assess interest in the article before it receives citations, to incorporate both academic and social metrics, and to allow a longitudinal view of the various signals. The metrics are grouped into categories [30.113]: viewed, saved, cited, recommended, and discussed. PLoS ALMs can be accessed through the PLoS API.

The major shortcoming of altmetric research at this point in time is that altmetrics provide counts of things without a clear meaning. Hopefully in the future we will have a better understanding, using qualitative methods together with quantitative ones. We must keep in mind that "not everything that counts can be counted, and not everything that can be counted counts" (a saying attributed to Albert Einstein).

30.14 A Wish-List for Future Data Collection from the Web

To sum up, I list the most important features that would enhance current data collection methods in my opinion:

- Transparency: This is a major problem, especially since the owners of the data are commercial companies. Examples of problems related to this aspect are: not knowing what is exactly covered by Google Scholar or the public Twitter API that samples somehow 10% of the data.
- Coverage: This issue was emphasized throughout the chapter, by comparing the number results reported by Google and Bing, low coverage of several altmetrics in general, and in the humanities especially (it is not easy or maybe even impossible to improve this situation, but I can still include it in my wish-list). The major point is that highly visible scholarly outputs are easily identifiable, and we need measures to assess items with medium or low visibility. We need measures that have good coverage if not of science as a whole, at least for the specific subject area being evaluated.
- Reliability and validity: In some cases there are huge fluctuations in the reported counts within a short period of time or inconsistencies in the reported number of results. This should be clearly avoided. Another aspect of reliability is continued availability of the information (this again is a wish that probably cannot be fulfilled, since platforms and tools change or even disappear over time). Validity is partially in the hands of the researcher. One has to be convinced that the data collected are the best possible data to answer the research questions.
- Search features and APIs: Please, bring the inlink operators back, and improve search on social media websites! Provide APIs to access data, and remove limitations (like maximum 1000 results displayed). Improve the precision and the recall of the retrieved data. This will result in less time spent on data cleansing. Improve handling of special characters that are problematic in several databases. Data quality of scholarly output can be improved by the use of unique identifiers (e.g., DOI, ORCiD).
- Provide access to web data and search engine logfiles: This again is easier said than done, because privacy issues are involved, and the interests of the commercial companies owning the data.
- Usage data: Publishers, aggregators, and repositories, please release more usage data.

Most of these points are not new, see for example [30.83, 114–117] for similar and additional recommendations.

To end the chapter on a positive note, research achievements in quantitative informetrics are quite remarkable, there are new developments and a growing interest. Researchers in the field have been able to overcome the difficulties and limitations in data collection and data cleansing by devising innovative methods and techniques. New and exciting research opportunities are ahead of us both in terms of techniques and methods (data science) and in terms of incentives (open science initiatives).

References

30.1 R. Caillieau: About WWW, J. Univers. Comput. Sci. **1**(4), 221–231 (1995)
30.2 Pew Research Center: World Wide Web timeline, http://www.pewinternet.org/2014/03/11/world-wide-web-timeline/ (2014)
30.3 K.A. Zimmermann: Internet History Timeline: ARPANET to the World Wide Web, https://www.livescience.com/20727-internet-history.html (2012)
30.4 R.R. Larson: Bibliometrics of the world wide web: An exploratory analysis of the intellectual structure of cyberspace. In: *Proc. 59th ASIS Annu. Meet., Baltimore* (1996)
30.5 Google: Refine web searches, https://support.google.com/websearch/answer/2466433?hl=en (2017)
30.6 T.C. Almind, P. Ingwersen: Informetric analyses on the world wide web: Methodological approaches to 'webometrics', J. Doc. **53**(4), 404–426 (1997)
30.7 P. Ingwersen: The calculation of web impact factors, J. Doc. **54**(2), 236–243 (1998)
30.8 A.G. Smith: A tale of two web spaces: Comparing sites using web impact factors, J. Doc. **55**(5), 577–592 (1999)
30.9 M. Thelwall: Web impact factors and search engine coverage, J. Doc. **56**(2), 185–189 (2000)
30.10 B. Cronin, H.W. Snyder, H. Rosenbaum, A. Martinson, E. Callahan: Invoked on the web, J. Am. Soc. Inf. Sci. **49**(14), 1319–1328 (1998)
30.11 R. Rousseau: Sitations: An exploratory study, Cybermetrics **1**(1), paper 1 (1997)
30.12 J. Bar-Ilan: The "mad cow disease", usenet newsgroups and bibliometric laws, Scientometrics **39**(1), 29–55 (1997)
30.13 J. Bar-Ilan: The mathematician, Paul Erdos (1913–1996) in the eyes of the internet, Scientometrics **43**(2), 257–267 (1998)
30.14 J. Bar-Ilan: On the overlap, the precision and estimated recall of search engines. A case study of the query "Erdos", Scientometrics **42**(2), 207–228 (1998)
30.15 J. Bar-Ilan, B. Peritz: The lifespan of "informetrics" on the web: An eight year study (1998–2006), Scientometrics **79**(1), 7–25 (2008)
30.16 M. Thelwall: Extracting accurate and complete results from search engines: Case study Windows Live, J. Am. Soc. Inf. Sci. Technol. **59**(1), 38–50 (2008)
30.17 J. Bar-Ilan, B.C. Peritz: A method for measuring the evolution of a topic on the web: The case of "informetrics", J. Am. Soc. Inf. Sci. Technol. **60**(9), 1730–1740 (2009)
30.18 W. Koehler: A longitudinal study of web pages continued: A consideration of document persistence, Inf. Res. **9**(2), 9–2 (2004)
30.19 D. Gomes, M.J. Silva: Modeling information persistence on the web. In: *Proc. 6th Int. Conf. Web Eng* (2006) pp. 193–200
30.20 R. Baeza-Yates, B. Poblete: Evolution of the chilean web structure composition. In: *Proc. IEEE/LEOS 3rd Int. Conf. Numer. Simul. Semicond. Optoelectron. Devices* (2003) pp. 11–13
30.21 H. Snyder, H. Rosenbaum: Can search engines be used as tools for web-link analysis? A critical view, J. Doc. **55**(4), 375–384 (1999)
30.22 J. Bar-Ilan: Search engine results over time: A case study on search engine stability, Cybermetrics **2/3**(1), paper 1 (1999)
30.23 J. Bar-Ilan: The web as an information source on informetrics? A content analysis, J. Am. Soc. Inf. Sci. **51**(5), 432–443 (2000)
30.24 W. Mettrop, P. Nieuwenhuysen: Internet search engines—fluctuations in document accessibility, J. Doc. **57**(5), 623–651 (2001)
30.25 J. Bar-Ilan: How much information do search engines disclose on the links to a web page? A longitudinal case study of the 'cybermetrics' home page, J. Inf. Sci. **28**(6), 455–466 (2002)
30.26 E. Sharp: The first page of Google, by the numbers, http://www.protofuse.com/blog/first-page-of-google-by-the-numbers/ (2014)
30.27 Wikipedia: Data cleansing, https://en.wikipedia.org/w/index.php?title=Data_cleansingoldid=771405405 (2017)
30.28 J. Bar-Ilan: Data collection methods on the web for infometric purposes—A review and analysis, Scientometrics **50**(1), 7–32 (2001)
30.29 M. Thelwall: Data cleansing and validation for multiple site link structure analysis. In: *Web Mining: Applications and Techniques*, ed. by A. Scime (IGI Global, Hershey 2005) pp. 208–227
30.30 M. Thelwall: Results from a web impact factor crawler, J. Doc. **57**(2), 177–191 (2001)
30.31 M. Thelwall: Extracting macroscopic information from web links, J. Am. Soc. Inf. Sci. Technol. **52**(13), 1157–1168 (2001)
30.32 M. Thelwall: A comparison of sources of links for academic web impact factor calculations, J. Doc. **58**(1), 66–78 (2002)
30.33 M. Thelwall: Conceptualizing documentation on the Web: An evaluation of different heuristic based models for counting links between university web sites, J. Assoc. Inf. Sci. Technol. **53**(12), 995–1005 (2002)
30.34 M. Thelwall, D. Wilkinson: Three target document range metrics for university web sites, J. Am. Soc. Inf. Sci. Technol. **54**(6), 490–497 (2003)
30.35 M. Thelwall: Evidence for the existence of geographic trends in university web site interlinking, J. Doc. **58**(5), 563–574 (2002)
30.36 M. Thelwall, R. Tang, L. Price: Linguistic patterns of academic web use in Western Europe, Scientometrics **56**(3), 417–432 (2003)
30.37 M. Thelwall, A. Smith: Interlinking between Asia-Pacific university web sites, Scientometrics **55**(3), 363–376 (2002)

30.38 M. Thelwall (Ed.): *Link Analysis: An Information Science Approach* (Elsevier, Amsterdam 2004)
30.39 M. Thelwall: Introduction to webometrics: Quantitative web research for the social sciences. In: *Synthesis Lectures on Information Concepts, Retrieval, and Services* (Morgan Claypool, San Rafael 2009)
30.40 L. Vaughan: Visualizing linguistic and cultural differences using web co-link data, J. Am. Soc. Inf. Sci. Technol. **57**(9), 1178–1193 (2006)
30.41 K.M. Kousha, M. Thelwall: Motivations for URL citations to open access LIS library and information science articles, Scientometrics **68**(3), 501–517 (2006)
30.42 P. Sud, M. Thelwall: Linked title mentions: A new automated link search candidate, Scientometrics **101**(3), 1831–1849 (2014)
30.43 H.J. Kim: Motivations for hyperlinking in scholarly electronic articles: A qualitative study, J. Am. Soc. Inf. Sci. **51**(10), 887–899 (2000)
30.44 D. Wilkinson, G. Harries, M. Thelwall, L. Price: Motivations for academic web site interlinking: Evidence for the web as a novel source of information on informal scholarly communication, J. Inf. Sci. **29**(1), 49–56 (2003)
30.45 J. Bar-Ilan: What do we know about links and linking? A framework for studying links in academic environments, Inf. Process. Manag. **41**(4), 973–986 (2005)
30.46 L. Vaughan: Exploring website features for business information, Scientometrics **61**(3), 466–477 (2004)
30.47 L. Vaughan, J. You: Comparing business competition positions based on web co-link data: The global market vs. the Chinese market, Scientometrics **68**(3), 611–628 (2006)
30.48 L. Vaughan, Y. Gao, M. Kipp: Why are hyperlinks to business websites created? A content analysis, Scientometrics **67**(2), 291–300 (2006)
30.49 L. Leydesdorff, M. Curran: Mapping university-industry-government relations on the internet: The construction of indicators for a knowledge-based economy, Cybermetrics **4**(1), 1–17 (2000)
30.50 D. Stuart, M. Thelwall: Investigating triple helix relationships using URL citations: A case study of the UK West Midlands automobile industry, Res. Eval. **15**(2), 97–106 (2006)
30.51 L. Vaughan, D. Shaw: Bibliographic and web citations: What is the difference?, J. Am. Soc. Inf. Sci. Technol. **54**(14), 1313–1322 (2003)
30.52 L. Vaughan, D. Shaw: Web citation data for impact assessment: A comparison of four science disciplines, J. Assoc. Inf. Sci. Technol. **56**(10), 1075–1087 (2005)
30.53 P. Jacsó: Google Scholar: The pros and the cons, Online Inf. Rev. **29**(2), 208–214 (2005)
30.54 P. Jacsó: Google scholar revisited, Online Inf. Rev. **32**(1), 102–114 (2008)
30.55 L.I. Meho, K. Yang: Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar, J. Am. Soc. Inf. Sci. Technol. **58**(13), 2105–2125 (2007)
30.56 C. Neuhaus, H.D. Daniel: Data sources for performing citation analysis: An overview, J. Doc. **64**(2), 193–210 (2008)
30.57 A.W.K. Harzing, R. Van der Wal: Google Scholar as a new source for citation analysis, Ethics Sci. Environ. Polit. **8**(1), 61–73 (2008)
30.58 M. Shultz: Comparing test searches in PubMed and Google Scholar, J. Med. Libr. Assoc. **95**(4), 442–445 (2007)
30.59 P. Jacsó: As we may search—comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases, Curr. Sci. **89**(9), 1537–1547 (2005)
30.60 K. Bauer, N. Bakkalbasi: An examination of citation counts in a new scholarly communication environment, D-Lib Magazine (2005), https://doi.org/10.1045/september2005-bauer
30.61 C. Neuhaus, E. Neuhaus, A. Asher, C. Wrede: The depth and breadth of Google Scholar: An empirical study, Portal **6**(2), 127–141 (2006)
30.62 M. Norris, C. Oppenheim: Comparing alternatives to the Web of Science for coverage of the social sciences' literature, J. Informetr. **1**(2), 161–169 (2007)
30.63 J.J. Meier, T.W. Conkling: Google Scholar's coverage of the engineering literature: An empirical study, J. Acad. Librariansh. **34**(3), 196–201 (2008)
30.64 K. Kousha, M. Thelwall: Google Scholar citations and Google web/URL citations: A multi discipline exploratory analysis, J. Am. Soc. Inf. Sci. Technol. **58**(7), 1055–1065 (2007)
30.65 A.W. Harzing: Publish or Perish, http://harzing.com/pop.gtm (2007)
30.66 D. Adams: Publish or Perish version 5, http://www.harzing.com/blog/2016/10/publish-or-perish-version-5 (2016)
30.67 A.W. Harzing, R. Van Der Wal: A Google Scholar h-index for journals: An alternative metric to measure journal impact in economics and business, J. Am. Soc. Inf. Sci. Technol. **60**(1), 41–46 (2009)
30.68 L. Bornmann, W. Marx, H. Schier, E. Rahm, A. Thor, H.D. Daniel: Convergent validity of bibliometric Google Scholar data in the field of chemistry—Citation counts for papers that were accepted by Angewandte Chemie International Edition or rejected but published elsewhere, using Google Scholar, Science Citation Index, Scopus, and Chemical Abstracts, J. Informetr. **3**(1), 27–35 (2009)
30.69 J.E. Hirsch: An index to quantify an individual's scientific research output, Proc. Natl. Acad. Sci. **102**(46), 16569–16572 (2005)
30.70 J. Bar-Ilan: Which h-index?—A comparison of WoS, Scopus and Google Scholar, Scientometrics **74**(2), 257–271 (2008)
30.71 L.I. Meho, Y. Rogers: Citation counting, citation ranking, and h-index of human computer interaction researchers: Comparison of Scopus and Web of Science, J. Am. Soc. Inf. Sci. Technol. **59**(11), 1711–1726 (2008)

30.72 A.W. Harzing: A preliminary test of Google Scholar as a source for citation data: A longitudinal study of Nobel prize winners, Scientometrics **94**(3), 1057–1075 (2013)
30.73 A.W. Harzing: A longitudinal study of google scholar coverage between 2012 and 2013, Scientometrics **98**(1), 565–575 (2014)
30.74 H.F. Moed, J. Bar-Ilan, G. Halevi: A new methodology for comparing Google Scholar and Scopus, J. Informetr. **10**(2), 533–551 (2016)
30.75 P. Jacsó: Deflated, inflated and phantom citation counts, Online Inf. Rev. **30**(3), 297–309 (2006)
30.76 J. Bar-Ilan: Citations to the "Introduction to Informetrics" indexed by WoS, Scopus and Google Scholar, Scientometrics **82**(3), 495–506 (2010)
30.77 E. Delgado López-Cózar, N. Robinson-García, D. Torres-Salinas: The Google Scholar experiment: How to index false papers and manipulate bibliometric indicators, J. Assoc. Inf. Sci. Technol. **65**(3), 446–454 (2014)
30.78 J. Pino-Díaz, E. Jiménez-Contreras, R. Ruíz-Baños, R. Bailón-Moreno: Strategic knowledge maps of the techno-scientific network (SK maps), J. Am. Soc. Inf. Sci. Technol. **63**(4), 796–804 (2012)
30.79 Google: Google books history, http://books.google.com/googlebooks/about/history.html
30.80 K. Kousha, M. Thelwall: Google book search: Citation analysis for social science and the humanities, J. Am. Soc. Inf. Sci. Technol. **60**(8), 1537–1549 (2009)
30.81 K. Kousha, M. Thelwall: An automatic method for extracting citations from Google Books, J. Assoc. Inf. Sci. Technol. **66**(2), 309–320 (2015)
30.82 DORA: San Francisco declaration on research assessment, http://www.ascb.org/files/SFDeclarationFINAL.pdf (2012)
30.83 D. Hicks, P. Wouters, L. Waltman, S. De Rijcke, I. Rafols: Bibliometrics: The Leiden Manifesto for research metrics, Nature **520**, 429–431 (2015)
30.84 E. Delgado López-Cózar, Á. Cabezas-Clavijo: Google Scholar Metrics: An unreliable tool for assessing scientific journals, http://digibug.ugr.es/bitstream/handle/10481/21540/GSM_castellano.pdf?sequence=6&isAllowed=y (2012)
30.85 E. Orduña-Malea, E.D. Delgado López-Cózar: Google scholar metrics evolution: An analysis according to languages, Scientometrics **98**(3), 2353–2367 (2014)
30.86 A.W. Harzing, S. Alakangas: Microsoft Academic: Is the phoenix getting wings?, Scientometrics **110**(1), 371–383 (2017)
30.87 P. Davis, M. Fromerth: Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles?, Scientometrics **71**(2), 203–215 (2007)
30.88 H.F. Moed: The effect of "open access" on citation impact: An analysis of ArXiv's condensed matter section, J. Am. Soc. Inf. Sci. Technol. **58**(13), 2047–2054 (2007)
30.89 V. Larivière, C.R. Sugimoto, B. Macaluso, S. Milojević, B. Cronin, M. Thelwall: ArXiv E-prints and the journal of record: An analysis of roles and relationships, J. Assoc. Inf. Sci. Technol. **65**(6), 1157–1169 (2014)
30.90 X. Li, M. Thelwall, K. Kousha: The role of arXiv, RePEc, SSRN and PMC in formal scholarly communication, Aslib J. Inf. Manag. **67**(6), 614–635 (2015)
30.91 J. Priem, D. Taraborelli, P. Groth, C. Neylon: Altmetrics: A manifesto, http://altmetrics.org/manifesto/ (2010)
30.92 S. Haustein, T.D. Bowman, R. Costas: Interpreting 'altmetrics': Viewing acts on social media through the lens of citation and social theories. In: *Theories of Informetrics and Scholarly Communication: A Festschrift in Honor of Blaise Cronin*, ed. by C.R. Sugimoto (De Gruyter, Berlin 2016) pp. 372–406
30.93 X. Li, M. Thelwall, D. Giustini: Validating online reference managers for scholarly impact measurement, Scientometrics **91**(2), 461–471 (2011)
30.94 J. Bar-Ilan, S. Haustein, I. Peters, J. Priem, H. Shema, J. Terliesner: Beyond citations: Scholars' visibility on the social web, https://arxiv.org/abs/1205.5611 (2012)
30.95 S. Haustein, V. Larivière, M. Thelwall, D. Amyot, I. Peters: Tweets vs. Mendeley readers: How do these two social media metrics differ?, IT-Inf. Technol. **56**(5), 207–215 (2014)
30.96 E. Mohammadi, M. Thelwall: Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows, J. Assoc. Inf. Sci. Technol. **65**(8), 1627–1638 (2014)
30.97 Z. Zahedi, R. Costas, P. Wouters: How well developed are altmetrics? A cross-disciplinary analysis of the presence of 'alternative metrics' in scientific publications, Scientometrics **101**(2), 1491–1513 (2014)
30.98 S. Haustein, I. Peters, J. Bar-Ilan, J. Priem, H. Shema, J. Terliesner: Coverage and adoption of altmetrics sources in the bibliometric community, Scientometrics **101**(2), 1145–1163 (2014)
30.99 M. Thelwall, S. Haustein, V. Larivière, C.R. Sugimoto: Do altmetrics work? Twitter and ten other social web services, PloS One **8**(5), e64841 (2013)
30.100 Altmetric Support: When did altmetric start tracking attention to each attention source?, https://help.altmetric.com/support/solutions/articles/6000136884-when-did-altmetric-start-tracking-attention-to-each-attention-source- (2017)
30.101 S. Haustein, I. Peters, C.R. Sugimoto, M. Thelwall, V. Larivière: Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature, J. Assoc. Inf. Sci. Technol. **65**(4), 656–669 (2014)
30.102 R. Costas, Z. Zahedi, P. Wouters: Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective, J. Assoc. Inf. Sci. Technol. **66**(10), 2003–2019 (2015)
30.103 H. Shema, J. Bar-Ilan, M. Thelwall: Do blog citations correlate with a higher number of future citations? Research blogs as a potential source

for alternative metrics, J. Assoc. Inf. Sci. Technol. **65**(5), 1018–1027 (2014)

30.104 L. Bornmann: Validity of altmetrics data for measuring societal impact: A study using data from Altmetric and F1000Prime, J. Informetr. **8**(4), 935–950 (2014)

30.105 E. Mohammadi, M. Thelwall: Assessing non-standard article impact using F1000 labels, Scientometrics **97**(2), 383–395 (2013)

30.106 P. Kraker, E. Lex: A critical look at the ResearchGate score as a measure of scientific reputation. In: *Proc. Quantif. Anal. Sch. Commun. Web Workshop, ASCW'15* (2015)

30.107 M. Thelwall, K. Kousha: ResearchGate: Disseminating, communicating, and measuring scholarship?, J. Assoc. Inf. Sci. Technol. **66**(5), 876–889 (2015)

30.108 M. Thelwall, K. Kousha: Academia.edu: Social network or academic network?, J. Assoc. Inf. Sci. Technol. **65**(4), 721–731 (2014)

30.109 Springer: Bookmetrix, http://www.springer.com/bookmetrix?SGWID=0-1773415-0-0-0 (2017)

30.110 J.C. Wallis, E. Rolando, C.L. Borgman: If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology, PloS One **8**(7), e67332 (2013)

30.111 European Commission: *Open Innovation, Open Science, Open to the World* (European Commission, Brussels 2016)

30.112 C. Neylon, S. Wu: Article-level metrics and the evolution of scientific impact, PLoS Biology **7**(11), e1000242 (2009)

30.113 PLOS: A comprehensive assessment of impact with article-level metrics (ALMs), https://www.plos.org/article-level-metrics

30.114 J. Bar-Ilan: Expectations versus reality–Search engine features needed for web research at mid 2005, Cybermetrics **9**, paper 2 (2005)

30.115 J. Wilsdon, L. Allen, E. Belfiore, P. Campbell, S. Curry, S. Hill, R. Jones, R. Kain, S. Kerridge, M. Thelwall, J. Tinkler, I. Viney, P. Wouters, J. Hill, B. Johnson: The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management, https://doi.org/10.13140/RG.2.1.4929.1363 (2015)

30.116 NISO: Altmetrics data quality code of conduct–Draft for public comment, http://www.niso.org/apps/group_public/document.php?document_id=16121wg_abbrev=altmetrics-quality (2016) NISO RP-25-201X-3

30.117 J. Wilsdon, J. Bar-Ilan, R. Frodeman, E. Lex, I. Peters, P. Wouters: *Next-Generation Metrics: Responsible Metrics and Evaluation for Open Science* (European Commission, Brussels 2017), https://ec.europa.eu/research/openscience/pdf/report.pdf

Judit Bar-Ilan

Information Science
Bar-Ilan University
Ramat Gan, Israel
judit.bar-ilan@biu.ac.il

Judit Bar-Ilan is a Professor at the Department of Information Science of Bar-Ilan University (Israel). She received her PhD in Computer Science from the Hebrew University of Jerusalem. She started her research on information science in the mid-1990s at the School of Library, Archive and Information Studies of the Hebrew University of Jerusalem. She moved to Bar-Ilan University in 2002. Her major areas of interest are informetrics and information retrieval.

31. Web Citation Indicators for Wider Impact Assessment of Articles

Kayvan Kousha

There are many ways in which academic articles can be used outside research contexts for teaching, culture, medical practice, business, policy making, or knowledge communication. Articles that have significant wider benefits may therefore be undervalued if they are assessed through conventional citation indicators and sources (e. g., the Web of Science (WoS) and Scopus). A range of online document genres and sources may help to evaluate these broader impacts of articles, including academic syllabi, textbooks, clinical trials or guidelines, patents, encyclopedia articles, and grey literature publications. Web citations can be used as a quantitative impact indicator for monitoring the wider impact of articles, especially in the arts, humanities, and social sciences where many research outputs have value beyond academia. This article reviews literature about the web citation analysis of articles and explains different methods to capture web citations from a range of online sources via commercial search engines. The applications and limitations of web citation analysis for wider impact assessment of articles are discussed, in addition to practical advice for data gathering. New web citation indicators can help research evaluation peer review and citation analysis by giving additional information about the wider benefits of published research when a type of impact (e. g., teaching, commercial, or clinical impact) is required to be assessed by authors, research funders, or evaluators in addition to their academic research impact.

31.1 Web as a Citation Source

Citation indicators derived from the Web of Science (WoS) and Scopus are often used to support research impact assessment for academic publications [31.1]. Nevertheless, formal citations from journal articles and conference papers are unlikely to be useful for monitoring the wider impacts of research. Alternative sources of citations may help research evaluations when the published research is cited in nonacademic publications (e. g., clinical trials or patents) or when an article's journal is not indexed by common citation databases (e. g., non-WoS journals). Alternative sources of citations may aid research evaluation exercises, where "all kinds of social, economic and cultural benefits and impacts beyond academia" need to be measured [31.2, p. 4]. Some scientometricians have recommended that indicators beyond bibliometrics [31.3, 4] and nonstan-

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_31

dard publications [31.5] are needed to examine all types of research impacts. This is because some research that has a significant intellectual impact may be undervalued if assessed through conventional citation indicators and databases. For instance, some scientific articles might be extensively used for teaching, clinical practice, technology transfer, policy making, or knowledge dissemination without being cited in academic publications.

There have been early initiatives to extract citations from web documents as part of the webometrics research area to develop alternative quantitative indicators for impact assessment (for reviews see: [31.6–8]). In a pioneering pair of studies, *Vaughan* and *Shaw* [31.9, 10] applied the term *web citation* to refer to citations that can be identified by searching for bibliographic information from an article, such as its title, author(s), journal name, or publication year, via commercial search engines. The assumption is that counting mentions of research in online sources can reflect the level of interest in scientific, scholarly related, educational, or cultural contexts. Web citations could be more useful in the arts and humanities than science, where scholarly outputs may have teaching, cultural, artistic, economic, or social influence rather than primarily academic impact, and because the coverage of current citation databases is insufficient for citation analyses of humanities research [31.11, 12]. Most previous web citation studies have examined online data sources to identify the broader benefits of academic outputs [31.13] and some web impact indicators have been proposed to assess wider benefits of funded scientific research [31.14–16].

The term *web citations* in this chapter refers to online mentions of scientific articles. Web citation indicators are different from social media metrics (known as altmetrics), which incorporate bookmarks, comments, views, downloads, and ratings from *social media* sites, such as Mendeley, CiteULike, LinkedIn, Twitter, and Facebook [31.17]. Both altmetrics and web citations are alternative indicators that may be called upon when the wider impacts of academic research are assessed. There have been several literature reviews and books about webometrics [31.6, 8, 18] and web citation indicators [31.13, 19, 20]. This chapter emphasizes web indicators for *articles* and their publishing journal, and summarizes what is known about new online sources and document genres (e. g., encyclopedia articles, grey literature publications, clinical sources, and news stories), search techniques, and disciplinary differences. Some of the web citation indicators introduced in this chapter, such as Google Books, syllabus and Wikipedia citations, can also be used for the impact assessment of books (Chap. 27).

31.2 Sources of Web Citations: Websites and Document Genres

This section introduces a range of types of website (e. g., blogs, Wikipedia) and document genres (e. g., syllabi, presentations) that may cite academic research. Their citations may be especially relevant when evaluators, funders, or scholars want to assess research impact beyond academia. Citations from these sources or genres may be identified with search engine queries or, in some cases, Application Programming Interface (API) queries, and the process can be automated to a large extent with free software.

There are four generic limitations with using the web for research evaluation purposes. First, not everything is online and so impact may occur without being recorded on the web. Second, it is likely that a biased fraction of impacts is recorded on the web. For example, impacts occurring in developing nations and for people that do not routinely publish online are likely to be underrepresented or absent. Third, commercial search engines do not index the entire web and so it is impossible to guarantee that all web citations to a document will be found. Thus, when using any type of web citation, researchers need to provide evidence that the results are meaningful. Finally, most web citation counts can be manipulated and so cannot be used if stakeholders are aware in advance that they will be used. This does not prevent formative applications and use for evaluations when the methods are unknown to stakeholders [31.21]. These limitations are discussed in more detail in the limitations section below.

This section lists the main types of web citation and a later section covers one of the ways (correlation tests) in which the validity of web indicators can be assessed.

31.2.1 Citations from the General Web

It is possible to identify citations to academic publications from the web using manual or automatic citation searches. Extra manual checks are often necessary to remove irrelevant matches, however.

General Web Citations

Commercial search engines (e. g., Google and Bing) can be used to manually or automatically locate mentions of academic articles from the web and this may reflect the overall interest in articles, at least as reflected online. *Bar-Ilan* [31.22] argued that the web is a free data source that can be used for bibliometric research. She searched the terms 'informetrics' and 'informetric' in major web search engines to identify the contexts in which they were used, finding that the web can locate relevant information not found in commercial bibliographic databases. *Vaughan* and *Shaw* [31.9] found that the web contains citations outside conventional citation indexes that can be identified online through searching with the bibliographic information of publications (see an example below). Nevertheless, the practical limitation of general web citation searches is that not all web citations are indicators of academic impact and web-extracted citations may be from library lists, tables of contents, current awareness databases, publishers' websites, or authors' CVs, which cannot be considered as evidence of scholarly impact. For example, there is evidence that only a third of general web citations from Google ($n = 854$) to library and information science journal articles were from references in other publications [31.9] and a follow-up study across four science subject areas (biology, genetics, medicine, and multidisciplinary science) found that 30% of web citations reflected intellectual impact, including citations from papers and from class reading lists [31.10]. Another study classified sources of 1577 web citations to open access journals from biology, physics, chemistry, and computing, finding a quarter of citations were from references of other publications [31.23]. These studies suggest that most general web citations could be mentioned for noncitation reasons (e.g., navigational), making this indicator less useful for intellectual impact assessment of articles. Nevertheless, general web citation searches could be more useful as evidence of impact for individual articles when searches are limited to a specific file type or website.

```
Jha "Chronic kidney disease:
Global dimension and
perspectives," Lancet 2013
```

URL Citations

URL citations of online publications have also been suggested as an alternative to web citations [31.24]. Although one advantage of URL citations over general web citations is that URLs are unique, giving less false matches than searching for articles by name, URLs of online articles might change over time and some journals may have multiple links to the same article in different formats (e.g., PDF or HTML). Moreover, some journals may use DOIs instead of URLs for directing users to articles and some authors may not mention URLs of online articles in their references, making URL citations less useful for impact assessment of research. Below is an example of a URL citation search for an article published in an open access journal, *Educational Technology & Society*.

```
https://www.j-ets.net/ETS/
journals/12_1/11.pdf
```

31.2.2 Citations from Specific Document Types

Web citation searches can be limited to specific document file types or specific internet domains to generate more targeted searches than general web citation queries.

Document Citations

Many papers, dissertations, and reports are available in Portable Document Format (PDF) or Microsoft Word format [31.25] and citations from these file formats may be more useful than general web citation searches for impact assessment [31.26]. Web citations from PDF, Microsoft Word (DOC and DOCX), or other document formats such as Rich Text Format (RTF) or PostScript (PS) can be counted using the *filetype:* search command in Google manual searches or automatically via the Bing API (see the examples below). For an inclusive web citation search from all document types, queries should be combined using the OR operator (Sect. 31.5), although the Google OR operator does not always work properly [31.27]. Document web citation searches are useful when evidence of the wider impact of research is required from online grey literature such as reports, white papers, or newsletters [31.28]. However, document web citation searches may retrieve many false matches from CVs in PDF or Microsoft Word format. It may also retrieve citations from the duplicate self-archived citing papers from authors' websites or different digital libraries (e.g., ResearchGate, Academia.edu, arXiv.org) and so manual checking of results may be necessary. Moreover, it seems unlikely that document web citation searches give more effective results than Google Scholar manual citation searches. In the absence of Google Scholar automatic searches for large-scale analyses, document citation web searches seem to be the only practical method to automatically estimate citations to articles from PDF or Microsoft Word format documents, unless Microsoft Academic can be used (see below).

```
Zhu "Prostate cancer in East
Asia: Evolving trend over the
last decade" 2015 filetype:pdf
```

```
Amini "Trends in hospital
volume and failure to rescue
for pancreatic surgery" 2015
filetype:doc
```

```
Shinozuka "Statistical analysis
of fragility curves" 2000
filetype:docx
```

Limiting a document type web citation search to a specific internet domain using the site: command is useful when web citations are required to be collected from academic (e. g., .edu, ac.uk, ac.jp), government (e. g., .gov), or national web domains (e. g., .fr, .de, or .ca), as shown in examples below (see also the grey literature citation section below):

```
Hainmueller Hopkins "Public
attitudes toward immigration"
"Annual Review of Political
Science" 2014 filetype:pdf
site:edu
```

```
Harper "Economic and social
implications of aging societies"
"Science" 2014 filetype:pdf
site:ca
```

31.2.3 Presentation Citations

There are many searchable academic presentations online in Microsoft PowerPoint format (.ppt and .pptx). Some scholars cite articles in their research or teaching presentation slides for conferences, seminars, or lectures. Citations from presentations could be more useful in conference-based fields, where presentations are important for disseminating scientific results [31.29]. Presentations for teaching or seminars may instead reflect the educational value of the cited research. Moreover, many scholars may not include references in their presentations, may not share their presentations online, or may publish them in PDF or Microsoft Word formats. Citations from presentations can be manually or automatically identified through searching for articles in presentation files and adding filetype:ppt or filetype:pptx to each query, as shown in the example below. It is important to combine web citation results from both ppt and pptx because they may give different matches (e. g., the second example below).

```
Liben-Nowell Kleinberg "The
link-prediction problem for
social networks" filetype:ppt
```

```
Kasimati "Economic aspects and
the Summer Olympics: A review of
related research" filetype:pptx
```

Another source of presentation citations is the SlideShare (slideshare.net) website, which is a specialized platform to share presentations (see an example below), although other scientific repositories, such as Figshare.org also include academic presentations [31.30]. A content analysis of SlideShare documents citing academic articles showed that most (64%) were academic publications (e. g., journal articles) rather than slides, making it less unique for presentation citation analysis [31.31].

```
Møller "Breast cancer survival
in England, Norway and Sweden:
a population-based comparison"
site:slideshare.net
```

Syllabus Citations

Course syllabi or reading lists are key documents for teaching and learning. University lecturers or teaching staff list academic publications in syllabi that are required or recommended for students to read. Mentions of publications in academic course syllabi or reading lists may reflect the educational benefits of the works listed [31.32]. Syllabus citations could help to estimate the educational value of teaching materials, such as textbooks and monographs. For instance, a study found that a third of about 14 000 Scopus-indexed monographs had at least one academic syllabus mention [31.33]. Syllabus citations can also be used to assess the teaching utility of academic articles. It is possible to locate syllabus citations through manual searches for academic papers in fully searchable syllabi that have been published online, although this requires extensive manual checking to filter out irrelevant matches. For instance, the bibliographic information of articles can be searched for in conjunction with the phrases `course description` or `syllabus` to identify syllabi that mention the articles searched for.

```
Harnad "Scholarly skywriting and
the prepublication continuum
of scientific inquiry" "course
description" site:edu
```

```
Gollin "The Lewis model:
A 60-year retrospective" syllabus
site:edu
```

To help identify academic research with high educational impact from university websites, a method has been developed to automatically search for mentions of articles in online academic course syllabi and filter out false matches, limiting search results to a list of over 24 600 university websites to help remove search matches originating from nonacademic sites. This method includes any search results from second-level academic domains, such as Australia (.edu.au/), Malaysia (.edu.my/), or Taiwan (.edu.tw/), as well as individual universities without specific academic domains, such as in Canada, France, and Germany [31.33].

There are also specialized educational websites that can be searched to assess the teaching benefits of research articles. One example in medical science is Radiopaedia.org, which provides radiology images and additional references related to them for educational purposes.

```
Angouras "Bovine aortic arch:
normal variant or a marker of
aortopathy?" Cardiology 2012
site:radiopaedia.org
```

31.2.4 Citations from High-Value Websites

Some important websites cite scientific articles for a specific purpose, including encyclopedia articles, clinical trials, or guidelines sites, and websites containing policy or planning documents, annual reports, science blogs or news. It is possible to use the `site:` command at the end of web citation queries to restrict the search results to a specific website and identify only citations from the given site.

Encyclopedia Citations

Citations from general or specialized encyclopedia entries to journal articles may reflect scientific, social, or cultural knowledge transfer to the wider public. Wikipedia citations, for instance, occur in scientific, educational, societal, and cultural contexts within encyclopedia entries (e.g., popular science, biographies, art, entertainment, history, and geography). To search for citations in Wikipedia articles with Google or Bing, bibliographic information from the articles should be combined with the wikipedia.org/wiki/ command (see an example below). Nevertheless, for a large-scale systematic analysis to identify Wikipedia citations Bing API searches should be used (see also the Web citation searching section). There is strong evidence that citations from Wikipedia articles are useful to assess the wider impacts of textbooks and monographs—a third of English monographs had at least one citation from Wikipedia [31.34]. Wikipedia citations can also be used for the wider impact assessment of articles, especially in the arts and humanities. As an example, the article below published in *Film History: An International Journal* had received over 50 Wikipedia citations without being cited in Scopus (as of April 2017).

```
Pierce "Forgotten faces: Why some
of our cinema heritage is part of
the public domain"
"Film History: An
International Journal" 2007
site:en.wikipedia.org
```

About 5% of the articles published by the Public Library of Science (PLoS) have been cited in Wikipedia [31.35] and a large-scale study of 302 328 Scopus articles published during 2005–2012 across multiple fields found that Wikipedia citations to articles were more frequent in History (11%) and in Music, Visual Arts and Performing Arts, Political Science (7%) than in Computer Science (1.4%) and Surgery (2%) [31.34].

Wikipedia citation searches can be restricted to entries in different languages by adding the language codes (see https://en.wikipedia.org/wiki/List_of_Wikipedias) to the site: command, such as for Spanish (es.wikipedia.org/wiki), Italian (it.wikipedia.org/wiki), or Persian (fa.wikipedia.org/wiki). For instance, the French article below had received two citations from French Wikipedia entries but none from English Wikipedia articles (see below). Nevertheless, because references in Wikipedia articles can be edited by any registered user, citations from Wikipedia can be manipulated or inflated [31.36]. Moreover, very few science articles are cited by Wikipedia, suggesting that Wikipedia citations could be most useful in the arts and humanities and perhaps some social science disciplines.

```
Messaoudi "Frontière du fractal
de Rauzy et système de numération
complexe" "Acta Arithmetica" 2000
site:fr.wikipedia.org/wiki/
```

In contrast to Wikipedia, there is a range of specialized and peer-reviewed online encyclopedias with quality-control mechanisms, such as Scholarpedia (scholarpedia.org) within the fields of mathematics and sciences (e.g., physical, biological, and behavioral sciences), Encyclopedia of Laser Physics and Technology (rp-photonics.com/encyclopedia.html), Stanford Encyclopedia of Philosophy (plato.stanford.edu), or The

A.D.A.M. Medical Encyclopedia (medlineplus.gov/ency) as shown in the examples below.

```
Ménard Bartelmann "Cosmological
information from quasar-galaxy
correlations induced by weak
lensing" "Astron Astrophys" 2002
site:scholarpedia.org
```

```
Jeong "Ytterbium-doped
large-core fiber laser with
1.36 kW continuous-wave
output power" "Opt. Express"
site:rp-photonics.com
```

```
Huttegger "Generic properties
of evolutionary games
and adaptationism" "The
Journal of Philosophy"
site:plato.stanford.edu
```

Medical Citations

Clinical trials and medical guidelines are important components of the process of improving healthcare in society in terms of diagnosis, health care and treatment of clinical problems [31.37]. It is possible to manually or automatically search for citations to medical research in some clinical trials and guideline websites (see also the Web citation searching section below). Web citation extraction methods have been used to count citations to medical journal articles from clinical guidelines from the *National Institute of Health and Clinical Excellence (NICE)* in the UK [31.38], clinical trials in ClinicalTrials.gov [31.39], and clinical drug information from the AHFS DI Essentials in Drugs.com [31.40]. These studies found that articles referenced in clinical publications tended to be more highly cited than comparable articles that were not cited in clinical documents. This suggests that articles with clinical practice value tend to attract more citations from other academic publications.

```
Niu "Percutaneous
cryoablation for stage IV
lung cancer: a retrospective
analysis" Cryobiology 2013
site:clinicaltrials.gov
```

```
Frans "Autism risk across
generations: a population-based
study of advancing
grandpaternal and paternal age"
site:.nice.org.uk
```

```
Granich "Multidrug resistance
among persons with tuberculosis
in California, 1994-2003" JAMA
site:drugs.com
```

There are other national clinical guidelines databases that may be used to assess the overall clinical influence or benefits of research such as Australian National Health and Medical Research Council clinical practice guidelines [31.41], Canadian Clinical Practice Guidelines [31.42], or Scottish Intercollegiate Guidelines Network [31.43].

Grey Literature Citations

Many national or international organizations produce grey literature in the form of annual reports, regulations, statistics, white papers, and government publications or policy documents. Citations from these can point to nonacademic types of impacts. Most of the grey literature is not published by academic or commercial publishers and hence not indexed in citation databases. Because of difficulties accessing grey literature, the *Grey Literature Report* (www.greylit.org/) from the New York Academy of Medicine has attempted to catalogue thousands of medical grey literature publications from many organizations, institutions, and medical centers around the world. Some of these sources are fully searchable by commercial search engines and can be found for web citation analysis. This includes in medicine and healthcare the World Health Organization (WHO), the Joint United Nations Programme on HIV/AIDS (UNAIDS), and the International Diabetes Federation, in environmental science the United Nations Environment Programme (UNEP) and the International Union for Conservation of Nature, in agriculture the Food and Agriculture Organization (FAO) and the International Food Policy Research Institute or in publications, reports, or statistics documents from the European Union. These organizations publish many publications, such as annual reports, white papers, regulations, statistics, guidelines, or government documents, that can be used for the wider impact assessment of research. Nevertheless, less than 0.5% of "policy-related documents" had at least one citation to WoS-indexed articles based on Altmetric data [31.44]. Below are some examples of web citation searches limited to PDF files in websites publishing grey literature.

```
Warren "Democracy and deceit.
Regulating appearances of
corruption" 2006 filetype:pdf
site:un.org
```

```
Rychetnik "Criteria for
evaluating evidence on public
health interventions" 2002
filetype:pdf site:who.int

Kumar "Antibiotic uptake by
plants from soil fertilized with
animal manure" site:eu
```

General web searches in huge websites such as WHO may also return results from CVs, library lists, or even formal publications such as published books, monographs, or journal articles, requiring extensive manual checks to identify correct grey literature citations. For instance, the query below retrieves one citation from an open access article published in Bulletin of the World Health Organization which is also indexed by both WoS and Scopus databases. Hence, it is important to exclude citations from these publications (e.g., http://www.who.int/bulletin) or to identify URL addresses associated with grey literature publications (e.g., http://apps.who.int/iris/) to avoid counting citations from journals or books instead of grey literature publications. However, some websites may use similar URLs for publishing their books, journals, or grey literature publications online, making web citation extraction more problematic.

```
Noor "A spatial national health
facility database for public
health sector planning in Kenya
in 2008" "Int J Health Geogr"
2009 site:who.int
```

ResearchGate Citations

ResearchGate can be used to identify early citations to articles from other uploaded publications. Each article in ResearchGate has a profile page, showing how many citations it has received from other documents in ResearchGate. DOIs of articles can be searched for in Bing or Google with a query that combines site:researchgate.net/publication with article information (e.g., title, publication year, journal name) to locate article pages in ResearchGate (see below). There is empirical evidence that ResearchGate citations are more numerous than both WoS and Scopus for recently published library and information science articles, but are less common than Google Scholar citations [31.45]. For instance, the query below locates five citations to an article published in 2016 ("Estimating open access mandate effectiveness: The MELIBEA score.") from uploaded preprints in ResearchGate without being cited either in WoS or Scopus (as of 24th of April 2017). Nevertheless, ResearchGate citations can be manipulated or inflated by uploading fake or nonpeer reviewed publications and an automatic method to extract ResearchGate citations for a large-scale study is likely to be problematic [31.45].

```
"DOI: 10.1002/asi.23601"
site:researchgate.net/publication
```

Blog Citations

Many science blogs disseminate and discuss scientific issues, such as those indexed in *ScienceBlogs.com, ResearchBlogging.org, blogs.nature.com*, and *blogs.plos.org*. Scholars may cite other research in their blog posts to support their claims, discussions, or reviews [31.46, 47]. There is evidence that science blog citations to other articles can show the wider impact of scientific articles [31.48, 49] and articles blogged in ResearchBlogging.org in the year of publication tended to receive more citations than other articles in the same journal [31.49]. Blog citations can be gathered by commercial search engines, although they may include links to other publications instead of cited references and sometimes mention research just for current awareness. Below is an example of a web citation search for a citation to an article from the ScienceBlogs website.

```
Nyffeler "An estimated 400-800
million tons of prey are annually
killed by the global spider
community" "The Science of
Nature" site:scienceblogs.com
```

31.2.5 Citations from Digital Libraries and Online Citation Indexes

Google Scholar Citations

Google Scholar systematically extracts citations from a huge number of scientific journals from many publishers, as well as conference papers, preprints and postprints of articles, dissertations, and other scholarly documents. Google Scholar citation counts are often substantially greater than those from conventional citation indexes [31.18, 50, 51], although this varies across subject areas [31.52]. Despite possible manipulation [31.53, 54], Google Scholar is useful for extracting and analyzing citations for impact assessment of articles [31.55]. Google Scholar is a better source of early citation impact evidence for new articles than WoS and Scopus because it takes longer for articles to be indexed by traditional citation indexes [31.45]. Although Google Scholar does not support automatic citation counting for a large-scale research evaluation of articles, the *Publish or Perish* software can be used for individual assessment of researchers and journals based on Google Scholar citation counts [31.56].

Google Books Citations

Books and monographs are important academic outputs and book citations can reflect a different type of impact compared with citations from journal articles [31.57]. Citations from scientific books to research articles presumably reflect a type of scientific impact. However, citations in books that are primarily written for teaching, such as textbooks for undergraduates, reflect educational impact instead. In some fields, such as history, arts, and literature, citations from books to academic articles may reflect cultural value. Because WoS and Scopus index a small number of books from selected academic publishers, citation indicators from them could be unrepresentative [31.58, 59]. Google Books seems to be a useful citation source because it contains millions of fully searchable books. In literature Google books citations to academic articles were twice as common as WoS citations but they were less numerous in library and information science (31%). They are more common in the arts, humanities, and social sciences than in the sciences [31.60]. Although it is possible to manually search for bibliographic information of articles in Google Books main search interface and to identify correct citations to articles (see an example below), this seems to be impractical for a large-scale impact assessment exercise. In response, a method has been constructed to automatically generate web citation queries for books (e.g., from WoS, Scopus, or other standard book sources) and search for citations to them via the Google Books API. Additional options are provided to filter out false matches from raw Google Books results such as book reviews, bibliographies, and book advertisement to identify correct citation matches with relatively high accuracy and coverage [31.13]. A Google Books API query may take the following form, which can also be submitted to the Google Books web interface.

```
Van Djik Hacker "The digital
divide as a complex and dynamic
phenomenon" "Information Society"
2003
```

Google Patents Citations

Citations from patents to academic publications may reflect the technological value of research [31.61, 62]. Google Patents covers a large collection of searchable patents from the multiple patent offices including the United States Patent and Trademark Office (USPTO) since 1790 and the European Patent Office (EPO) since 1978 as well as patents from the World Intellectual Property Organization (WIPO), Canada, China, and Germany. Hence, the Google Patents website can be a useful source of citations to scientific articles in patent references. To generate Google Patent searches, article bibliographic information should be combined with the `site:google.com/patents` command to restrict the results to the Google Patents website (see example below). Although Google Patents citation searches can be conducted manually through Google, the Bing API can be used to automatically submit the queries instead. In some fields with a patenting culture, such as Biomedical Engineering, Biotechnology, and Pharmacology & Pharmaceutics, patent citations are numerous enough to help assess the commercial value of academic articles [31.63].

```
Lagasse "Purified hematopoietic
stem cells can differentiate
into hepatocytes in vivo"
"Nature Medicine" 2000
site:google.com/patents
```

Microsoft Academic Citations

Like Google Scholar, Microsoft Academic indexes millions of scholarly publications, providing citation counts for them. Unlike Google Scholar, the new version of Microsoft Academic allows a limited number of automatic searches per month which is useful for large-scale bibliometric analysis of articles [31.64]. Several recent studies have shown that Microsoft Academic finds similar numbers of citations or slightly more than Scopus and WoS [31.65–67] and finds substantially more citations for recently-published or in press articles [31.68, 69]. Because Microsoft Academic has faster citation indexing than conventional citation indexes and wide coverage of e-print archives [31.64, 70], it seems to be the most useful automatic tool for monitoring the early citation impact of in press or recently published articles.

Other Digital Libraries

Open access repositories such as *CogPrints* or *bioRxiv* can also be used to extract citations from preprints, postprints, or working papers. However, documents in these archives may also be indexed by Google Scholar or by conventional citation indexes after they have been formally published and hence web citations from these digital libraries may not be very useful for wider impact assessment of articles.

```
Hill "Mating Games: The Evolution
of Human Mating Transactions"
"Behavioral Ecology" 2004
filetype:pdf site:cogprints.org

Vasquez "Short term serum
pharmacokinetics of diammine
silver fluoride after oral
application" "BMC Oral Health"
2012 site:biorxiv.org
```

31.3 Web Citation Indicators for Journals

There have been efforts to use web citation indicators for journals rather than individual articles, motivated by journal impact factors. Web citation indicators can be derived for journals by calculating the average number of web citations to articles published in a journal during a year, irrespective of the publication year of the citing web page. This could then be compared with the Journal Impact Factor (JIF) or Scopus journal metrics. There is some empirical evidence about the value of web citation indicators for journal impact assessment. The average number of general web citations to journal articles highly correlates with JIFs for library and information science [31.10] and four science fields (Vaughan & Shaw, 2005), suggesting that web citations could also be useful for assessing the impact of journals. There is also evidence that counts of citations from Wikipedia to journal articles significantly correlate with JIFs [31.71].

It is reasonable to examine the web citation impact of journals based upon relevant online sources that cite them. For example, articles published in the journal *Expert Opinion on Therapeutic Patents* (Taylor & Francis Group) or *Drug Development Research* (Wiley) may attract many citations from patents (both with over 100 citations). Similarly, articles in clinical medicine journals such as *Clinical Infectious Diseases* or *Journal of Clinical Oncology* may receive many citations from clinical trials or guidelines and articles published in *Journal of Medical Biography* or *British Journal for the History of Science* may be widely used in encyclopedia articles to provide historical background information.

Online news stories are another source of web citations, where research findings with general interest or relevance for the public may be reported [31.72, 73]. Since journalists rarely mention article titles and authors in their news stories, web citation analysis seems to be not useful for individual article impact assessment purposes. Nevertheless, several studies have shown that most news stories name journals as their sources [31.74–76]. Mentions of journals can be captured through searching for journal names in searchable news websites (e. g., bbc.co.uk, cnn.com, and reuters.com) or newspapers (e. g., The Guardian or The Independent). For short generic journal names (e. g., Science and Nature) the term *journal* needs to be added before or after their names to reduce false matches during web citation searches. Nevertheless, software is required to extract journal citations from the matching news stories [31.77].

```
"Journal of Archaeological
Science" site:theguardian.com

"journal Nature" site:bbc.co.uk

"Lancet journal" site:reuters.com
```

There are also specialized news sources like *Medscape* (medscape.com), *Environmental News Network* (*enn.com*), or *SpaceDaily* (spacedaily.com) that can be used for the impact assessment of journals in different disciplines (see examples below).

```
"British Journal of Cancer"
site:medscape.com

"journal Nature Climate Change"
site:.enn.com

"journal Physical Review Letters"
site:spacedaily.com
```


31.4 Types of Web Citation Impacts

Some web citation indicators can reflect a different type of impact to that of traditional citations such as educational, cultural, clinical treatment, commercial, knowledge dissemination, public awareness, and policy. These types of intellectual impacts are helpful for assessing the wider influence or benefits of academic articles, although collecting web citations is generally much more difficult than using traditional citation databases (see also below limitations of web citation analysis). Scholars may believe that their academic outputs have wider impacts rather than, or in addition to, research impact. For instance, some academic articles may be frequently mentioned in academic course syllabi or textbooks for their educational benefits or receive relatively many citations in medical clinical records because of their value in healthcare and patient treatments. Other articles may be cited in patents by inventors for their commercial or technological values or cited in annual reports or regulations by organizations or governments to support their arguments about policy changes. Table 31.1 summarizes different types of intellectual impacts from web citation sources and their application for the impact assessment of articles.

Table 31.1 Type of web citation indicators from different online sources and their limitations and potential use for impact assessment of articles

Web source	Web citation source	Impact type	Limitations	Overall usefulness (Poor/Fair/Good)	Citation overlap with WoS/Scopus
General web	All academic or nonacademic web sites and publications	General web impact	• Possible manipulation of citations • Many false matches and navigational citations • Needs extensive manual checks • Includes diverse impact types	Poor	To some extent
Document files	Online PDF or Microsoft Word (.doc or docx) documents	General web impact	• Possible manipulation of citations • Needs manual check of results • Less effective than Google Scholar citations	Fair	To some extent
Presentations	Online presentations (.ppt or pptx)	Educational; Scholarly	• Possible manipulation of citations • Needs manual checks of results • Missing presentations in other formats (e. g., PDF), many academic presentations are not searchable by search engines	Fair	Not at all
Syllabi	Online searchable academic course syllabi or reading lists	Educational	• Needs manual checks of results, although software can help filtering nonsyllabus matches • Many academic syllabi are not searchable online	Good	Not at all
Encyclopedias	Online searchable encyclopedias	Scholarly; educational; cultural; arts and history	• Possible manipulation of citations for editable encyclopedias (e. g., Wikipedia), although this is less likely for peer-reviewed online encyclopedias (e. g., Stanford Encyclopedia of Philosophy)	Good	Not at all
Clinical practices	Online searchable clinical guidelines or trails	Health care and treatment	• Many clinical practise documents are not searchable online	Good	Not at all
Grey literature	Digitized reports, statistics, regulations, guidelines, and policy documents	Policy making; educational; economic; legislative	• May need manual checks of results • Searches may include false matches from online CVs or journals or bulletins published by organizations • Many grey literature publications are not searchable online	Good	Not at all
ResearchGate	Uploaded publications (e. g., preprints or postprints of articles)	Scholarly; early impact	• Possible manipulation of citations • Automatic citation searches of large numbers of articles may not be possible	Fair	To a great extent
Blogs	Science blog posts	Scholarly; public engagement	• Possible manipulation of citations • May need manual checks of results	Fair	Not at all
Google Scholar	Citations from a large number of scholarly documents	Scholarly; Early impact	• Possible manipulation of citations • Needs manual checks of results • Automatic citation searches of many articles may not possible	Good	To a great extent
Google Books	Citations from a large number of digitized books	Scholarly; educational	• Needs manual checks of results based on live searches, although software can filter out false matches	Good	Very little
Google Patents	Citations from a large number of digitized patents	Commercial; technological	• Needs manual checks of results • Most articles are not cited by patents and patent citations could be helpful in some fields (e. g., biomedical engineering)	Good	Not at all

31.5 Web Citation Searching

In this section, heuristics for conducting effective web searches are discussed. These involve counting the number of times that an article is cited in online documents. Searches need to have an acceptable level of accuracy and coverage to be useful for any given task. This includes 1) generating appropriate web citation queries from article bibliographic information and 2) submitting the queries to a search engine to locate citations.

Although it is possible to manually generate queries and use the Google or Bing web search interfaces to identify web citations for a few articles, for a large data set free software can be used to automate web citation searches. Web citation queries can be generated and searched automatically through the free Webometric Analyst software (http://lexiurl.wlv.ac.uk, see also [31.20]). This software can construct web citation queries for different web sources (e. g., Google Patents, Wikipedia, syllabi, PowerPoint, Blogs, or document file types) from a set of WoS or Scopus data. To generate queries using Webometric Analyst, article bibliographic information should be in plain text (tab delimited) Unicode format. The software automatically searches the queries generated in Bing through the Bing web search API. Users need to register for the Bing Search API key and enter it in Webometric Analyst prior to running searches, allowing 1000 free searches per month for up to three months. For a large-scale research evaluation exercise involving many articles and multiple queries, users need to purchase extra API transactions to automatically collect the web citation data (https://azure.microsoft.com/en-us/pricing/details/cognitive-services/search-api/web/). Webometric Analyst also supports automatic query building and locating citations in digitized books via Google Books API [31.78] in addition to the automatic searching and filtering options to identify academic syllabus citations [31.33].

Web citation queries can be generated by combining the last name of the first author, phrase searches for the article title and journal name, and the publication year. Any bibliographic database listing articles in a standard format, such as WoS, Scopus Medline, Compendex, or ERIC, can be used to generate web citation queries in this way. It is recommended that each query contains the last names of the first (up to) three authors in order to reduce false matches. For an article with more than three authors, sometimes the first author's names plus 'et al.' might be mentioned in the reference lists and so including all authors' names in the queries may considerably reduce the number of correct web citation matches. Below is an example of a journal article cited in a Wikipedia entry mentioning only the first author of a publication with six authors.

```
White, Howard D., et al.
"Libcitations: A measure
for comparative assessment
of book publications in the
humanities and social sciences."
Journal of the American Society
for Information Science and
Technology 60.6 (2009):
1083-1096.
```

For articles with very short or general titles (e. g., *When less is more* or *Developing an HIV vaccine*), it is necessary to add extra bibliographic information to the queries such as journal names to prevent retrieving too many false matches. Nevertheless, a practical problem is that depending on academic citation styles, journal names may be mentioned in full (e. g., *Journal of Cancer Clinical Trials* in APA (American Psychological Association) or Harvard citation styles) or in abbreviated format (e. g., *J Cancer Clin Trials* in Vancouver or NLM (National Library of Medicine) styles) in the reference lists of publications. Hence, journal names in the web citation queries should be used carefully. Abbreviated journal titles are more frequently used within medical or physical sciences than social sciences. For instance, the Google search of query *LANCET INFECT DIS site:clinicaltrials.gov* retrieved 204 results from online clinical trials, whereas no result was found using the full journal name *('Lancet Infectious Diseases" site:clinicaltrials.gov*). Different abbreviations might be used for a single journal in the cited references of publications such as *N. Engl. J. Med.* (ISO Abbreviation) or *NEW ENGL J MED* (JCR Abbreviation) for *The New England Journal of Medicine*. Thus, for a comprehensive web citation search of articles with short or common titles, it is important to conduct two or three separate queries at the same time covering full and abbreviated journal names and to combine the results using the OR operator or the vertical bar "|" in Webometric Analyst Bing automatic search (see an example below and [31.20]). The vertical bar in Webometric Analyst instructs it to submit two searches and combine the results, avoiding potential problems due to inadequate processing of OR queries by search engines. One advantage of searching abbreviated journal names (Drug Dev. Res. or DRUG DEVELOP RES) over full journal names with very generic titles ("Drug Development Research") is that they return much fewer false matches, particularly from large fully searchable databases such as Google Books or Google Patents, especially in subject areas where abbreviated journal names are used in the references (e. g., NLM, AMA, or Vancouver). Both WoS and Scopus provide full and abbreviated journal names in their export files that can be used for this purpose.

```
Streit Matsumoto "African
trypanosomiasis" "New England
Journal of Medicine" 2016 |
Streit Matsumoto "African
trypanosomiasis" "N. Engl. J.
Med." 2016 | Streit Matsumoto
"African trypanosomiasis" "NEW
ENGL J MED" 2016
```

Although the most common method to locate web citations from web documents is through manual or automatic searches in commercial search engines, the contents of some websites may not be well indexed by search engines and hence web citation extraction might not be useful. In this case, it might be possible to extract citations from web documents through a free crawler such as SocSciBot, (http://socscibot.wlv.ac.uk). This needs a list of URLs and another specific program to capture citations from downloaded pages such as from ResearchGate, ClinicalTrials.gov, or Drugs.com (see the Services tab in the free Webometric Analyst software). Several studies have used this method for citation analysis instead of searching for articles via Bing or Google [31.39, 40, 45].

31.6 Correlations Between Web Citation Indicators and Citation Counts for Academic Articles

There is empirical evidence that some web citation indicators for academic articles positively correlate with counts of citations from traditional citation databases. Correlation tests have been used to assess the degree of similarity between web citation indicators and traditional citations and to examine whether a similar type of citation impact from different scholarly sources (e. g., books, patents, encyclopedia articles, and course syllabi) can be observed from either WoS or Scopus citations. A strong correlation would suggest that the two data sources reflect similar types of academic research impact, while a weak correlation suggests that they reflect different types of intellectual impact or that at least one is affected by random or spurious influences. Several early investigations found that general web or URL citations to journal articles correlate significantly and positively with counts of citations from journal articles across multiple subject areas, suggesting that both could be in some way reflecting a similar type of intellectual impact. The average number of Google web citations to journal articles correlated significantly with JIFs for library and information science in 1998 (0.59) and 1999 (0.43) [31.10]. Stronger associations were found in biology (0.72), genetics (0.68), medicine (0.74), and multidisciplinary journals (0.79) [31.10]. Another study also found significant Spearman correlations between WoS and Google web citation counts for open access journal articles across four science and four social science disciplines (except psychology), ranging from 0.70 in computer science and 0.52 in sociology to 0.26 in biology and 0.29 in chemistry [31.79]. These investigations gave initial evidence that web citations could be useful for assessing the impact of journal articles.

Google Scholar citations have the highest positive correlations with WoS or Scopus citation counts and Google Scholar commonly retrieves more citations to individual articles than WoS and Scopus [31.50, 79–82]. A study of 2675 recently published library and information science articles (2016–March 2017) found higher correlations between Google Scholar and ResearchGate citations (0.732) than between Google Scholar with either WoS (0.582) or Scopus (0.624), suggesting that both online databases reflect early citations in a similar way [31.45].

Spearman correlations between WoS and Google Books citations are high in computer science (0.71; perhaps because conference proceedings were indexed as books), philosophy (0.65), linguistics (0.61), psychology (0.59), and social science (0.48) [31.60]. Nevertheless, correlations between conventional citations and other web citation indicators are low, and so not all web indicators are useful for scholarly impact evaluation, although they may be useful to identify wider impacts of research. For instance, Spearmen correlations between syllabus mentions and WoS citations to 1987 journal articles in library and information science are low but significant (0.23) [31.32] and weak but positive significant correlations were found between Google Patent and Scopus citations to over 322 000 articles in 16 science and engineering fields. The correlations were highest in Biomedical Engineering (0.36) and lowest in Industrial & Manufacturing Engineering (0.05) [31.63]. There were also significant, but low, Spearman correlations between Scopus and Wikipedia citations to over 300 000 articles across multiple fields (e. g., ranging from 0.105 in social sciences to 0.179 in science subject areas), suggesting that Wikipedia citations do not directly reflect academic research impact [31.34] and similarly an investigation found a weak correlation (0.201) between blog citations and WoS citations to 13 300 articles in medicine and biological sciences [31.48].

31.7 Limitations of Web Citation Analysis

Using the web for informetric research is much more challenging than traditional citation analysis [31.83]. For web citation analysis, software is usually needed for data collection and running many searches via Bing API keys can be expensive. Although Google Scholar generally locates more citations than Scopus or WoS, it does not allow automatic data collection. Microsoft Academic seems to be a suitable alternative, however. Coverage of searches, incorrect matches, and possible manipulation of online citations are other limitations for web impact assessment of articles.

31.7.1 Coverage

Search engines only report web citations from the part of the web that they crawl [31.84, 85]. For example, many academic course reading lists or syllabi are not indexed by commercial search engines. Moreover, Google probably covers more of the web than Bing [31.86], which is a limitation for web citation analyses with Bing, especially for PDF or PowerPoint files. Some websites may only be indexed by one search engine. For instance, only Google returns web citation results from The Munich Personal RePEc Archive (https://mpra.ub.uni-muenchen.de/) which includes about 40 000 preprints and postprints of articles and working papers in the field of economics (see an example below). Another limitation of web citation searches, and especially for journal-level impact assessment, is that commercial search engines such as Google and Bing return a maximum of 1000 results per query, although this seems to be less problematic for monitoring the web citation impact of individual articles.

```
Bonanno "Intensity of competition
and the choice between product
and process innovations"
site:mpra.ub.uni-muenchen.de
```

31.7.2 False Matches

Web citation searches may return false matches and this is more common for articles with general or short titles (e. g., "*Introduction to American literature*" or "*Negative certainty*"), especially in conjunction with common author names (e. g., Schmidt, Jones, or Yan) and generic journal names (e. g., *Nature*, *Science*, *Alcohol*, or *Appetite*). Sometimes there is no practical method to exclude irrelevant citation matches from web citation results. Reporting absolute citation counts for such cases could be problematic and hence it is necessary to manually check a sample of web citation results, especially those with extreme web citation counts. For instance, searching a query `Pennisi "Making waves" Science 2004` in Google retrieves several false citation matches in PDF files in the sense of not citing the correct article because the citation information of the article had been captured elsewhere within the references (see Fig. 31.1).

Most unwanted matches from web citation searches in the general web are likely to be due to capturing mentions of articles in noncitation contexts such as online CVs, journal tables of contents, library lists, and publisher websites [31.10, 23]. For Google Books citations, manual searches may give many irrelevant results and it is necessary to conduct manual checking to identify correct citation matches. For instance, some citations in digitized books in Google Books may occur in book reviews, bibliographies, advertisements, and abstracting and indexing volumes. Similarly, it is very difficult to identify correct academic syllabus citations through manual checking of web citation results for large research evaluation exercises, although this is possible for small numbers of articles from individual authors or departments. Methods have been designed to automatically filter false matches from both the Google Books [31.78] and academic syllabi citation searches [31.33] with relatively high accuracy (over 90%). However, the filtering mechanisms used are only applicable to English language books and syllabi, and may miss many relevant results in other languages, requiring additional filtering rules to identify correct web citations for different languages.

31.7.3 Duplicate Results

Web citation searches may return duplicate matches from multiple versions of the same publication, such

Pennisi, E. (2004). Bonemaking protein shapes beaks of Darwin's finches. *Science* **305**, 1383.
Phippard, D. J., Weber-Hall, S. J., Sharpe, P. T., Naylor, M. S., Jayatalake, H., Maas. R., Woo, I., Roberts-Ciark, D., Francis-West, P. H., Liu, Y. H., Maxson, R., Hili, R. E., and Dale, T. C. (1996). Regulation of Msx-1. Msx-2, Bmp-2 and Bmp-4 during foetal and postnatal mammary gland development. *Development* **122**, 2729–2737.
Plikus, M., and Chuong, C.-M. (2004). Making waves with hairs. *J. Invest. Dermatol.* **122**, VII–IX.

Fig. 31.1 An example of a false citation match from Google web citation search result as PDF or HTML versions of articles or preprints/postprints of articles shared by different authors in different platforms such as their CVs, academic social websites or repositories. Duplicate citation counting may also occur during Google Patent citation searches from different patent versions (e. g., submitted, revised, and accepted, see [31.63]) as well as through Wikipedia citation searches from articles in different languages with the same contents (e. g., entries translated from one language to other languages, see [31.34]). Hence, it is important to filter out identical citing documents based on their titles. Moreover, sometimes grey literature citation searches based on limiting the results to a specific website may also retrieve citations from journals or bulletins published by governments or organizations that are indexed by conventional citation indexes. Similarly, most citations from SlideShare documents seem to be published journal articles or conference papers rather than presentations. Duplicate citations with traditional citation indexes should not be considered as evidence of wider research impact, although it might be very difficult to identify these cases during systematic data collection.

31.7.4 Manipulation

Web citations can be manipulated or inflated by users to increase citation statistics by generating fake documents with references to be indexed by Google, Google Scholar or Bing or by manipulating citations in Wikipedia entries, or other public web documents. Nevertheless, the risk is much lower for Google Books, Google Patents, academic course syllabi and clinical trials from approved sources because apparently they cannot easily be manipulated by users.

31.7.5 Diverse Types of Impacts

In contrast to the scholarly impact reflected by citations from refereed journal articles, some web citation indicators reflect types of impacts that are difficult to characterize or that may depend on the discipline of the cited article. Google Books or Wikipedia citations to articles may represent scientific, teaching, or cultural impacts. For example, the article "Outcome of symptoms of dizziness in a general practice community sample" published in the journal *Family Practice* has been mostly cited in medical textbooks (e. g., Oxford Textbook of Primary Medical Care) in addition to some scientific books (e. g., "Neuro-Otology"). Citations to articles from patents may also be used for different reasons, such as the commercial benefits of research or other impacts in theoretical contexts (e. g., "Polycaprolactone (PCL) is a polymer used for implantable/injectable drug delivery systems for medical implants"). Similarly, articles may be cited in presentations for research (conference slides) or teaching (lecture slides). To assess the type of impacts in a given context, a content analysis of the citations or citing sources should be carried out. This can be very time consuming, expensive, and the results can also be somewhat subjective.

31.8 Conclusions

Citation analysis is commonly used to evaluate the research impact of scientific articles based on counting citations mainly from journal articles and conference papers indexed by WoS or Scopus. In this chapter, a range of web citation indicators has been introduced to help monitor the impact of academic articles from online books, patents, clinical trials, encyclopedia articles, grey literature publications, and academic syllabi. These web indicators may give some useful quantitative information about the wider impacts of articles outside conventional citation indicators or assist human judgements about the contributions of articles. Web citation indicators can also provide additional information for research funders and evaluators to identify successful research based upon educational, cultural, commercial, or societal impacts.

Google Scholar seems to be useful for the early scholarly impact assessment of articles across different subject areas from a wide range of publications in different languages. Google Books citations from textbooks, monographs, edited books, and reports seems to be useful in the arts, humanities, and social sciences. Google Patents citations can help to identify articles with commercially or technologically relevant benefits in some areas, especially in engineering fields. Microsoft Academic seems to be the most useful web-based automatic tool for assessing the citation impact of articles for large-scale analyses, especially when it is important to know the impacts of preprints or recently published articles (e. g., 1–2 years after publication or peer review). Citations from academic course reading lists can reveal the educational value of published articles and citations from encyclopedia articles can reflect the value of articles for knowledge communication to the wider public. In medical and health care sciences, clinical publications can be used to assess clinical benefits of articles such as for the treatment of patients. There is a range of digitized grey literature publications such as annual reports, guidelines, regulations, statistics, and government publications that may be used as citation evidence for the wider impacts of articles on national or international policies or strategies. News stories citing journals could also be useful to examine which scientific journals or academic subjects are most frequently of interest to the public.

There is evidence that different web citation indicators significantly correlate (but with low to medium strength) with counts of citations from journal articles at the article level and in some cases at the journal level, indicating that web citations are not random and have similarities with traditional citation counts. Google Scholar and Google Books citations have the highest correlations with traditional citation counts because they contain citations from academic publications. There are low but significant correlations between WoS/Scopus citations to articles and other web citation indicators such as syllabus and encyclopedia citations, which reflect wider benefits of research such as teaching or cultural value. Hence, some web citation indicators may be more useful in the arts and humanities and social science, where broader publication types and scholarly related outputs such as textbooks, encyclopedia articles, and course reading lists can also help to identify the value of published articles.

Although some web citation indicators may contain errors and can be manipulated, they may still be used to support peer review or WoS/Scopus citation counts, especially in subject areas where peer review is more subjective (e. g., humanities) and WoS/Scopus citations are rare or not useful to assess wider impact of academic articles. In science and medicine, some authors may claim that a particular type of web indicator such as citations in digitized books, clinical trials, or patents, more realistically reflect the value of their research than WoS/Scopus citations. Hence, it is important to identify web indicators that reflect broader impacts of research. For instance, articles about novel solar cell or smartphone technologies are more likely to receive citations in patents than in course syllabi. Similarly, articles with clinical contributions may attract more citations in clinical documents than general encyclopedias or articles about the history of films, music, or literature may relatively receive more citations in educational sources than policy documents. Moreover, some web citation indicators should be used cautiously for recently published articles because in some subject areas such as the arts and humanities and social sciences research takes a longer time to be cited than science and medicine. This also depends on citing sources like patents and books where they need more time to be published than articles and blog posts.

Finally, there are many subject-oriented websites that can be used for wider impact assessment of articles in specific academic subject areas in addition to the general web sources introduced in this chapter. Hence, it is important to identify potential relevant websites, such as digital libraries, repositories, online encyclopedias, and key organizations publishing grey literature publications in the relevant fields. If this could be achieved in accordance with the nature of different subject areas, then web citation indicators could more comprehensively reflect the wider impacts of academic publications.

References

31.1 H.F. Moed: *Citation Analysis in Research Evaluation* (Springer, New York 2005)

31.2 REF: Decisions on assessing research impact, http://www.ref.ac.uk/media/ref/content/pub/decisionsonassessingresearchimpact/01_11.pdf (2011)

31.3 T.C. Almind, P. Ingwersen: Informetric analyses on the World Wide Web: Methodological approaches to "Webometrics", J. Doc. **53**(4), 404–426 (1997)

31.4 B. Cronin: Bibliometrics and beyond: some thoughts on web-based citation analysis, J. Inf. Sci. **27**(1), 1–7 (2001)

31.5 G. Lewison: Beyond outputs: New measures of biomedical research impact, Aslib Proc. **55**(1/2), 32–42 (2003)

31.6 M. Thelwall, L. Vaughan, L. Björneborn: Webometrics, Annu. Rev. Inf. Sci. Technol. **39**, 81–135 (2005)

31.7 J. Bar-Ilan: Informetrics at the beginning of the 21st century–A review, J. Informetr. **2**(1), 1–52 (2008)

31.8 M. Thelwall: Introduction to Webometrics: Quantitative web research for the social sciences. In: *Synthesis Lectures on Information Concepts, Retrieval, and Services* (Morgan and Claypool Publishers, San Francisco 2009) pp. 1–116

31.9 L. Vaughan, D. Shaw: Bibliographic and Web citations: What is the difference?, J. Am. Soc. Inf. Sci. Technol. **54**(14), 1313–1324 (2003)

31.10 L. Vaughan, D. Shaw: Web citation data for impact assessment: A comparison of four science disciplines, J. Am. Soc. Inf. Sci. Technol. **56**(10), 1075–1087 (2005)

31.11 D. Hicks: The difficulty of achieving full coverage of international social science literature and the bibliometric consequences, Scientometrics **44**(2), 193–215 (1999)

31.12 A. Nederhof: Bibliometric monitoring of research performance in the social sciences and the humanities: A review, Scientometrics **66**(1), 81–100 (2006)

31.13 K. Kousha, M. Thelwall: Web impact metrics for research assessment. In: *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*, ed. by B. Croninand, C.R. Sugimoto (MIT Press, Cambridge 2014) pp. 289–306

31.14 The ACUMEN Consortium: Guidelines for Good Evaluation Practice with the ACUMEN Portfolio (2014) http://research-acumen.eu/wp-content/uploads/D6.14-Good-Evaluation-Practices.pdf

31.15 J. Wilsdon, L. Allen, E. Belfiore, P. Campbell, S. Curry, S. Hill, R. Jones, R. Kain, S. Kerridge, M. Thelwall, J. Tinkler: *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management* (Higher Education Funding Council for England, Stoke Gifford 2015)

31.16 M. Thelwall, K. Kousha, A. Dinsmore, K. Dolby: Alternative metric indicators for funding scheme evaluations, Aslib J. Inf. Manag. **68**(1), 2–18 (2016)

31.17 J. Priem, D. Taraborelli, P. Groth, C. Neylon: Altmetrics: A manifesto, http://altmetrics.org/manifesto/ (2010)

31.18 J. Bar-Ilan: Which h-index? – A comparison of WoS, Scopus and Google Scholar, Scientometrics **74**(2), 257–271 (2008)

31.19 M. Thelwall, K. Kousha: Web indicators for research evaluation, Part 1: Citations and links to academic articles from the web, Prof. Inf. **24**(5), 587–606 (2015)

31.20 M. Thelwall: *Web Indicators for Research Evaluation: A Practical Guide* (Morgan & Claypool, San Rafael 2017)

31.21 P. Wouters, R. Costas: *Users, Narcissism and Control: Tracking the Impact of Scholarly Publications in the 21st Century* (SURFfoundation, Utrecht 2012) pp. 847–857

31.22 J. Bar-Ilan: The Web as information source on informetrics? A content analysis, J. Am. Soc. Inf. Sci. **51**(5), 432–443 (2000)

31.23 K. Kousha, M. Thelwall: How is science cited on the Web? A classification of Google unique web citations, J. Am. Soc. Inf. Sci. Technol. **58**(11), 1631–1644 (2007)

31.24 K. Kousha, M. Thelwall: Motivations for URL citations to open access library and information science articles, Scientometrics **68**(3), 501–517 (2006)

31.25 E. Jepsen, P. Seiden, P. Ingwersen, L. Björneborn, P. Borlund: Characteristics of scientific Web publications: Preliminary data gathering and analysis, J. Am. Soc. Inf. Sci. Technol. **55**(14), 1239–1249 (2004)

31.26 K. Kousha: Characteristics of open access web citation networks: A multidisciplinary study, Aslib Proc. **61**(4), 394–406 (2009)

31.27 J. Bar-Ilan: Expectations versus reality – Search engine features needed for Web research at mid 2005, Cybermetrics **9**(2) (2005), https://dialnet.unirioja.es/descarga/articulo/1292632.pdf

31.28 D. Wilkinson, P. Sud, M. Thelwall: Substance without citation: Evaluating the online impact of grey literature, Scientometrics **98**(2), 797–806 (2014)

31.29 M. Thelwall, K. Kousha: Online presentations as a source of scientific impact?: An analysis of PowerPoint files citing academic journals, J. Am. Soc. Inf. Sci. Technol. **59**(5), 805–815 (2008)

31.30 M. Thelwall, K. Kousha: Figshare: A universal repository for academic resource sharing?, Online Inf. Rev. **40**(3), 333–346 (2016)

31.31 M. Thelwall, K. Kousha: SlideShare presentations, citations, users and trends: A professional site with academic and educational uses, J. Assoc. Inf. Sci. Technol. **68**(8), 1989–2003 (2017)

31.32 K. Kousha, M. Thelwall: Assessing the impact of research on teaching: An automatic analysis of online syllabuses in science and social sciences, J. Am. Soc. Inf. Sci. Technol. **59**(13), 2060–2069 (2008)

31.33 K. Kousha, M. Thelwall: An automatic method for assessing the teaching impact of books from online academic syllabi, J. Assoc. Inf. Sci. Technol. **67**(12), 2993–3007 (2015)

31.34 K. Kousha, M. Thelwall: Are Wikipedia citations important evidence of the impact of scholarly articles and books?, J. Assoc. Inf. Sci. Technol. **68**(3), 762–779 (2017)

31.35 J. Priem, H.A. Piwowar, B.M. Hemminger: Altmetrics in the wild: using social media to explore scholarly impact, http://arxiv.org/abs/1203.4745v1 (2012)

31.36 M.D. Bould, E.S. Hladkowicz, A.E. Pigford, L. Ufholz, T. Postonogova, E. Shin, S. Boet: References that anyone can edit: Review of Wikipedia citations in peer reviewed health science literature, BMJ (Online) **348**, g1585 (2014)

31.37 D. Kryl, L. Allen, K. Dolby, B. Sherbon, I. Viney: Tracking the impact of research on policy and practice: Investigating the feasibility of using citations in clinical guidelines for research evaluation, BMJ Open **2**(2), e000897 (2012)

31.38 M. Thelwall, N. Maflahi: Guideline references and academic citations as evidence of the clinical value of health research, J. Assoc. Inf. Sci. Technol. **67**(4), 960–966 (2016)

31.39 M. Thelwall, K. Kousha: Are citations from clinical trials evidence of higher impact research? An analysis of ClinicalTrials.gov, Scientometrics **109**(2), 1341–1351 (2016)

31.40 M. Thelwall, K. Kousha, M. Abdoli: Is medical research informing professional practice more highly cited? Evidence from AHFS DI Essentials in Drugs.com, Scientometrics **112**(1), 509–527 (2017)

31.41 Australian Government: National Health and Medical Research Council, https://www.nhmrc.gov.au

31.42 Canadian Medical Association: CPG Infobase: Clinical Practice Guidelines, https://www.cma.ca/En/Pages/clinical-practice-guidelines.aspx

31.43 Scottish Intercollegiate Guidelines Network: Current Guidelines, https://www.sign.ac.uk/our-guidelines.html

31.44 R. Haunschild, L. Bornmann: How many scientific papers are mentioned in policy-related doc-

uments? An empirical investigation using Web of Science and Altmetric data, Scientometrics **110**(3), 1209–1216 (2017)
31.45 M. Thelwall, K. Kousha: ResearchGate versus Google Scholar: Which finds more early citations?, Scientometrics **112**(2), 1125–1131 (2017)
31.46 I. Mewburn, P. Thomson: Why do academics blog? An analysis of audiences, purposes and challenges, Stud. High. Educ. **38**(8), 1105–1119 (2013)
31.47 H. Shema, J. Bar-Ilan, M. Thelwall: Research blogs and the discussion of scholarly information, PLoS ONE **7**(5), e35869 (2012)
31.48 M. Thelwall, S. Haustein, V. Larivière, C. Sugimoto: Do altmetrics work? Twitter and ten other candidates, PLoS ONE **8**(5), e64841 (2013)
31.49 H. Shema, J. Bar-Ilan, M. Thelwall: Do blog citations correlate with a higher number of future citations? Research blogs as a potential source for alternative metrics, J. Assoc. Inf. Sci. Technol. **65**(5), 1018–1027 (2014)
31.50 L.I. Meho, K. Yang: Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar, J. Am. Soc. Inf. Sci. Technol. **58**(13), 2105–2125 (2007)
31.51 J.C.F. de Winter, A.A. Zadpoor, D. Dodou: The expansion of Google Scholar versus Web of Science: A longitudinal study, Scientometrics **98**(2), 1547–1565 (2014)
31.52 H.F. Moed, J. Bar-Ilan, G. Halevi: A new methodology for comparing Google Scholar and Scopus, J. Informetr. **10**(2), 533–551 (2016)
31.53 J. Beel, B. Gipp: Academic search engine spam and Google Scholar's resilience against it, J. Electron. Publ. (2010), https://doi.org/10.3998/3336451.0013.305
31.54 E.D. López-Cózar, N. Robinson-García, D. Torres-Salinas: The Google Scholar experiment: How to index false papers and manipulate bibliometric indicators, J. Assoc. Inf. Sci. Technol. **65**(3), 446–454 (2014)
31.55 M. Khabsa, C.L. Giles: The number of scholarly documents on the public web, PLoS ONE **9**(5), e93949 (2014)
31.56 A.W.K. Harzing, R. van der Wal: Google Scholar as a new source for citation analysis, Ethics Sci. Environ. Politics **8**(1), 61–73 (2008)
31.57 B. Cronin, H. Snyder, H. Atkins: Comparative citation rankings of authors in monographic and journal literature: A study of sociology, J. Doc. **53**(3), 263–273 (1997)
31.58 J. Gorraiz, P.J. Purnell, W. Glänzel: Opportunities for and limitations of the book citation index, J. Am. Soc. Inf. Sci. Technol. **64**(7), 1388–1398 (2013)
31.59 D. Torres-Salinas, R. Rodríguez-Sánchez, N. Robinson-García, J. Fdez-Valdivia, J.A. García: Mapping citation patterns of book chapters in the Book Citation Index, J. Informetr. **7**(2), 412–424 (2013)
31.60 K. Kousha, M. Thelwall: Google book search: Citation analysis for social science and the humanities, J. Am. Soc. Inf. Sci. Technol. **60**(8), 1537–1549 (2009)
31.61 M. Huang, W. Huang, D. Chen: Technological impact factor: An indicator to measure the impact of academic publications on practical innovation, J. Informetr. **8**(1), 241–251 (2014)
31.62 Y. Liaw, T. Chan, C. Fan, C. Chiang: Can the technological impact of academic journals be evaluated? The practice of non-patent reference (NPR) analysis, Scientometrics **101**(1), 17–37 (2014)
31.63 K. Kousha, M. Thelwall: Patent citation analysis with Google, J. Assoc. Inf. Sci. Technol. **68**(1), 48–61 (2017)
31.64 S.E. Hug, M. Ochsner, M.P. Brändle: Citation analysis with Microsoft Academic, Scientometrics **111**(1), 371–378 (2017)
31.65 A.W. Harzing, S. Alakangas: Microsoft Academic: Is the phoenix getting wings?, Scientometrics **110**(1), 371–383 (2017)
31.66 A.W. Harzing, S. Alakangas: Microsoft Academic is one year old: The Phoenix is ready to leave the nest, Scientometrics **112**(3), 1887–1894 (2017)
31.67 M. Thelwall: Microsoft Academic: A multidisciplinary comparison of citation counts with Scopus and Mendeley for 29 journals, J. Informetr. **11**(4), 1201–1212 (2017)
31.68 M. Thelwall: Does Microsoft Academic find early citations?, Scientometrics **114**(1), 325–334 (2018)
31.69 K. Kousha, M. Thelwall: Can Microsoft Academic assess the early citation impact of in press articles? A multi-discipline exploratory analysis, J. Informetr. **12**(1), 287–298 (2018)
31.70 M. Thelwall: Can Microsoft Academic be used for citation analysis of preprint archives?, The case of the Social Science Research Network, Scientometrics **115**(2), 913–928 (2018)
31.71 F.A. Nielsen: Scientific citations in Wikipedia, First Monday (2007), http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1997/1872
31.72 M.F. Weigold: Communicating science: A review of the literature, Sci. Commun. **23**(2), 164–193 (2001)
31.73 A. Dudo: Scientists, the media, and the public communication of science, Sociol. Compass **9**(9), 761–775 (2015)
31.74 V. Entwistle: Reporting research in medical journals and newspapers, BMJ **310**(6984), 920 (1995)
31.75 C. Bartlett, J. Sterne, M. Egger: What is newsworthy? Longitudinal study of the reporting of medical research in two British newspapers, Br. Med. J. **325**(7355), 81–84 (2002)
31.76 G. Lewison, S. Tootell, P. Roe, R. Sullivan: How do the media report cancer research? A study of the UK's BBC website, Br. J. Cancer **99**(4), 569 (2008)
31.77 K. Kousha, M. Thelwall: An automatic method to identify citations to journals in news stories: A case study of the UK newspapers citing Web of Science journals, Scientometrics (submitted)
31.78 K. Kousha, M. Thelwall: An automatic method for extracting citations from Google Books, J. Assoc. Inf. Sci. Technol. **66**(2), 309–320 (2015)
31.79 K. Kousha, M. Thelwall: Google Scholar citations and Google Web/URL citations: A multi-discipline exploratory analysis, J. Am. Soc. Inf. Sci. Technol. **58**(7), 1055–1065 (2007)

31.80 A.V. Kulkarni, B. Aziz, I. Shams, J.W. Busse: Comparisons of citations in Web of Science, Scopus, and Google Scholar for articles published in general medical journals, J. Am. Med. Assoc. **302**(10), 1092–1096 (2009)
31.81 J. Mingers, E.A.E.C.G. Lipitakis: Counting the citations: A comparison of Web of Science and Google Scholar in the field of business and management, Scientometrics **85**(2), 613–625 (2010)
31.82 E.M. Lasda Bergman: Finding citations to social work literature: The relative benefits of using Web of Science, Scopus, or Google Scholar, J. Acad. Librar. **38**(6), 370–379 (2012)
31.83 J. Bar-Ilan: Data collection methods on the web for informetric purposes – A review and analysis, Scientometrics **50**(1), 7–32 (2001)
31.84 S. Lawrence, C.L. Giles: Accessibility of information on the web, Nature **400**(6740), 107–109 (1999)
31.85 W. Mettrop, P. Nieuwenhuysen: Internet search engines – fluctuations in document accessibility, J. Doc. **57**(5), 623–651 (2001)
31.86 D. Lewandowski: Evaluating the retrieval effectiveness of web search engines using a representative query sample, J. Assoc. Inf. Sci. Technol. **66**(9), 1763–1775 (2015)

Kayvan Kousha

Statistical Cybermetrics Research Group
University of Wolverhampton
Wolverhampton, UK
k.kousha@wlv.ac.uk

Dr. Kayvan Kousha is a researcher in the Statistical Cybermetrics Research Group at the University of Wolverhampton, UK. His research contribution has been mainly in the development of novel methods for extracting and analyzing web citations for wider impact assessment of research. He has published over 50 journal articles and 20 conference papers about online citations such as digitized books, patents, encyclopedia articles, clinical documents, and course syllabi.

32. Usage Bibliometrics as a Tool to Measure Research Activity

Edwin A. Henneken, Michael J. Kurtz

Measures for research activity and impact have become integral ingredients in the assessment of a wide range of entities (individual researchers, organizations, institutions, regions, disciplines). Traditional bibliometric indicators, like publication- and citation-based indicators, provide an essential part of this picture, but cannot describe the complete picture. Since reading scholarly publications is an essential part of the research lifecycle, it is only natural to introduce measures for this activity in attempts to quantify the efficiency, productivity and impact of an entity. Citations and reads are significantly different signals, so taken together, they provide a more complete picture of research activity. Most scholarly publications are now accessed online, making the study of reads and read patterns possible. Clickstream logs allow us to follow information access by the entire research community in real time. Publication and citation datasets just reflect activity by authors. In addition, download statistics, derived from these clickstreams, will help us identify publications with significant impact, but which do not attract many citations. Clickstream signals are arguably more complex than, say, citation signals. For one, they are a superposition of different classes of readers. Systematic downloads by crawlers also contaminate the signal, as does random browsing behavior. We will discuss the complexities associated with clickstream data and how, with proper filtering, statistically significant relations and conclusions can be inferred from download statistics. We will describe how download statistics can be used to describe research activity at different levels of aggregation, ranging from organizations to countries. These statistics show a strong correlation with socioeconomic indicators, like the gross domestic product (GDP). A comparison will be made with traditional bibliometric indicators. Since we will be using clickstream data from the Astrophysics Data System (ADS), we will argue that astronomy is representative for more general trends.

32.1 Previous Studies and Scope

The standard indicators to measure the quality and quantity of scholarly research are funds expended, number of papers published, and number of citations to those papers and measures derived from these citations. Since the turn of the century a fourth key indicator for research assessment has arisen: measures of the use of the (now almost exclusively) digital research documents. The concept of *research documents* represents a more general class of expressions of research than *scholarly publications*, because it contains anything that may get published during research. This study is confined to scholarly publications, that is articles in scholarly journals. Scholarly research can be represented by identifying a *research cycle*; research consists of activity expended in the various stages of this cycle (illustrated in Fig. 32.1). We assume that taking the publication stage of the research cycle as a proxy will sufficiently represent *research activity*. Also, from a practical point of view, the publication stage is the only stage in this cycle that allows for clearly quantifiable metrics. Therefore we focus on usage of scholarly publications.

The use of this usage information is in its infancy. The leading assessments of the quality and quantity of

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_32

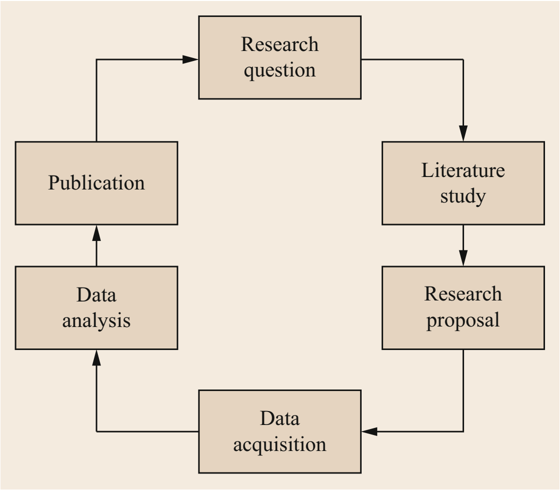


Fig. 32.1 Graphical representation of the research cycle

research on a country basis (Science & Engineering Indicator [32.1], for a given year), a university basis (Times Higher Education [32.2] or ARWU [32.3] ranking, for a given year), or on a journal basis (Impact Factor [32.4], Eigenfactor [32.5, 6] or Source Normalized Impact per Paper [32.7], for a given year) do not use usage information. The only widespread use of digital download records is by librarians making purchase decisions, aided by the COUNTER [32.8] standard for a given year, continuing their practice from the print era.

Since the first obsolescence function based on digital downloads was published 20 years ago [32.9] there has been an avalanche of work on the nature of digital download information. The review by *Kurtz* and *Bollen* [32.10] contains 171 references; more recent work with extensive discussions and bibliographies include [32.11–18]. Download information is now commonly found on article pages on journal websites and on various aggregator websites. Perhaps the most influential of these are the article download counts for a given year by the SSRN (formerly known as Social Science Research Network).

Before usage information can be widely used in making practical decisions some obvious problems need to be addressed. Just as the most popular wine, novel, or politician is not necessarily the best one, pure usage counts of scholarly article downloads may not be an accurate measure of research activity.

Users of scholarly research articles can be crudely divided into four categories [32.10, 11]: researchers, practitioners, students, and the general public. It is not uncommon that researchers represent just a small fraction of the total use. The number of healthcare practitioners dwarfs the number of medical researchers, for example. Students [32.19] use the scholarly literature very differently from researchers. In astronomy the number of interested lay people exceeds the number of research astronomers by a factor of perhaps 10 000; the number of serious amateur astronomers is about 100 times the number of researchers, while the number of citizen scientists on a single astronomy project, Galaxy Zoo, is a factor of ten larger than the world total of researchers in astronomy [32.20]. All these groups can and do access and read the scholarly literature. However, just the user class of researchers contribute to usage corresponding to research activity.

In this paper we use download measures to assess the quantity and to some degree quality of astronomy research, with the expectation that the techniques will also prove useful in other fields. We are aided in this by the fact that there are few professional astronomers who are not researchers; there is no such thing as applied astronomy. This simplifies the task compared with a field such as medicine, where a majority of professionals are not researchers. There are two different aspects to download measures: who is downloading and what is downloaded. Both are quantifiable signals and we will use both of them in our analysis. In order to get meaningful results, measures for both aspects are essential.

To some degree, research activity is related to economic trends. In previous publications we have explored relationships between economic indicators and research-related indicators [32.21–24]. This present paper substantially expands on the download analysis in these publications and shows that the data used in this analysis are consistent with the conclusions from these earlier studies.

32.2 Definition of Terminology

Use and usage are terms that seem to have a range of different meanings in the literature on bibliometrics. Other terms often encountered are *hits*, *clicks*, *reads* and *downloads*. Used as a measure, usage reflects aspects of user interaction with a digital service, in this case the Smithsonian/NASA Astrophysics Data System (ADS). During a session, users typically access different types of information by clicking on the appropriate links. In this study, we restrict ourselves to *downloads*, which is defined as clicking on a link that will take a user to the full text version of an article (either stored locally on an ADS server or stored externally). In the presentation and discussion of our results, we will use the terms *downloads* and *usage*, but they will both refer to the act of getting to the full article text.

The terms listed in Table 32.1 will be used frequently in the description of data (among others) and are particularly important with respect to creating and interpreting data. Hence, it's important to define them.

The data in this study have been derived from the usage logs of the ADS. These usage logs contain the user interactions with the user interface, recording the nature of the interaction (the type of information being requested), the time of this interaction (the date and time in Eastern Standard Time, EST), the identifier associated with the user and the IP address of the source requesting the information. The originating country of a request is derived from the hostname of the request, which in turn is derived from the IP address. In the mapping from hostname to country, we associated the top level domains *.net*, *.edu*, *.gov* and *.mil* with the USA.

When logging user interactions with the ADS, we attempt to filter out robots based on our knowledge of so-called *User Agents* and origin IP addresses of these requests. Since these robot requests are mostly for metadata, missed robot requests in our filtering will not contaminate our data, because we are focusing our study on downloads. Figure 32.1 shows the number of users in various categories over the period of our analysis (2005–2015). The line representing the *remainder* in this diagram represents situations like computers in libraries. In an earlier study [32.24] we observed that the median of the number of reads for frequent users is fairly constant at a value of about 21 reads per month. Since this number includes more than downloads, we decided to take 100 downloads per year as the lower boundary for the frequency interval associated with *frequent users*. Usage data for frequent users suggests a reads to download ratio of between 2 and 3. A second argument indicating that our choice of the definition for frequent users is meaningful is illustrated by the number of frequent users when restricted to downloads of publications from the main astronomy journals (line with open triangles in Fig. 32.2). This restriction will underestimate the number of research astronomers, so it is a lower limit. In the period 2005–2015, the total number of members of the International Astronomical Union (IAU) grows from around 9000 in 2005 to around 12 000 in 2015, represented by the dashed line. We only have actual IAU membership numbers for the period of 2008–2015. The fact that this line is bounded by the two lines representing both types of frequent users is a strong indication that our definition of frequent user can be interpreted as a representative definition.

Figure 32.1 also characterizes the type of signal we are looking at in this study. We will be looking at the access of full texts (*downloads*) by frequent users, which is roughly two orders of magnitude smaller than the total access. Also, as shown in an earlier study (Fig. 3 in [32.24]), the readership pattern of frequent users is significantly different from that of incidental users. Figure 32.2 indicates that the class of frequent users of the ADS exceeds the number of professional astronomers, the excess mostly likely consisting of physicists and engineers. Since most professional astronomers are using the ADS on a daily basis, it makes sense to focus on the field of astronomy in our analysis, because this will result in signals that can be regarded as truly representative for the entire field.

Why do we focus on the downloads by frequent users? Essentially because all authors are ADS users, but not all ADS users are authors. A comparison of the obsolescence functions of reads and citations makes this abundantly clear. Figure 32.3 (taken from [32.24]) illustrates this fact. From the ADS usage logs for January and February of 2008, we display the usage by frequent users of the ADS with those coming in from Google Scholar (representing incidental users) and compare these signals with citation rates and total citations. This analysis has been restricted to publications from the main astronomy journals. The fact that the relative use for frequent users follows the citation rate, rather than total citations, as function of publication year illustrates why we should focus on usage by frequent users. Citations, by their very nature, are deliberate and by using usage by frequent users, we expect to distill a signal that is expected to represent actions by authors with the least amount of noise. Note that

Table 32.1 Definition of frequently used terms

Term	Definition
Entity	A group of people generating publications, which can be identified through queries in the ADS, and whose interaction with the ADS can be easily identified in the usage logs. The following examples come to mind: country, institute/organization
Frequent user	This is an ADS user identified from the usage logs as one with at least 100, but no more than 1000, downloads per year. The reason for the upper limit is the fact that there are download sessions that cannot be associated with one particular individual (like computers in libraries)
Main astronomy journals	The Astrophysical Journal, The Astrophysical Journal Letters, The Astrophysical Journal Supplement Series, The Astronomical Journal, Monthly Notices of the Royal Astronomical Society, Astronomy & Astrophysics

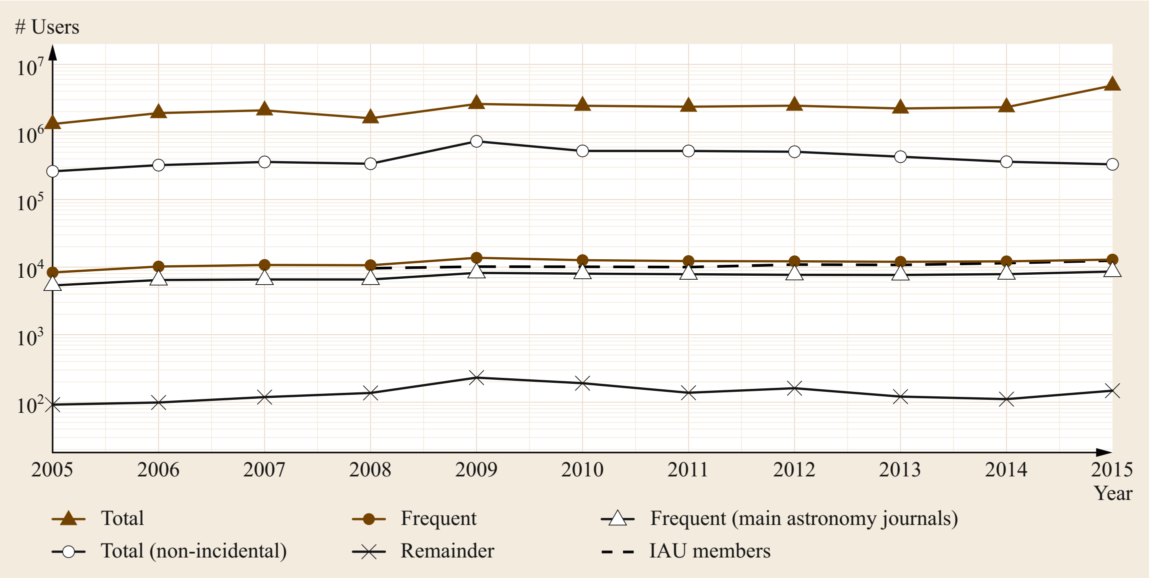


Fig. 32.2 The number of ADS users during the period of analysis, based on the number of yearly downloads. The line with *solid triangles* shows the total number of users for each year, the line with *open circles* shows the total number of users who downloaded at least one publication (this excludes those users who just look at an abstract), the line with *solid circles* shows the number of users who accessed the full text of between 100 and 1000 publications per year and the line with *crosses* represents the remainder. The line with *open triangles* shows the number of users who accessed the full text in the main astronomy journals of between 100 and 1000 publications per year. The *dashed line* represents the total number of IAU members

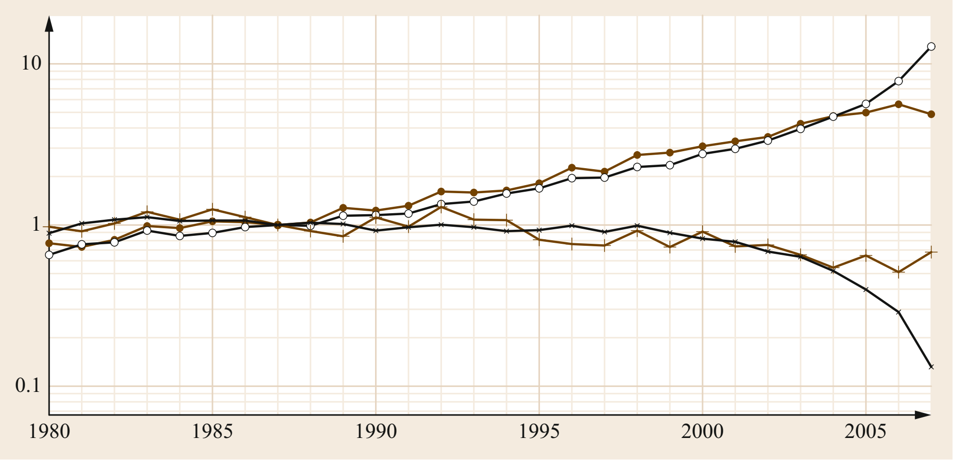


Fig. 32.3 Comparison of readership patterns from ADS and Google Scholar queries, as observed in the ADS access logs. The line marked with *open circles* shows the readership use by people using the ADS search engine. The line marked with + corresponds with the readership use by people who used the Google Scholar engine. The line marked with *solid circles* shows the citation rate of the articles, while the line marked with x represents their total number of citations

an additional interpretation of Fig. 32.3 is that Google Scholar does not provide results that researchers are looking for.

The other component of research activity considered in our study consists of the scholarly publications generated by entities analyzed. For each year in the period analyzed, a bibliography is compiled for any given entity using the query capabilities of the beta version of the ADS, codenamed *Bumblebee* (http://ui.adsabs.harvard.edu). The main reason for using this version of

Table 32.2 Overview of publication numbers for the start and end of the period, for a number of entities

Entity	Total		Refereed		Main astronomy	
	2005	2015	2005	2015	2005	2015
Yale	677	1130	502	945	79	194
Princeton	1479	2338	1157	1603	195	264
CfA	2541	1950	713	1061	497	824
NOAO	283	251	232	246	218	238
Canada	5785	9055	4209	6494	395	893
The Netherlands	3444	6120	2697	4169	442	841
Argentina	899	1376	793	1062	85	139
India	3484	9278	3190	6927	153	339

the ADS, rather than the older version, ADS *Classic*, is the fact that it supports a rich query language, allowing queries that involve affiliation information. With a couple of exceptions, the bibliographies were determined with the following query:

```
aff: "<affiliation string>"
year: 2005-2015
```

We used the application programming interface (API) of ADS *Bumblebee* (https://github.com/adsabs/adsabs-dev-api) to generate the bibliographies with metadata identifying the publication, the affiliation of the authors and the refereed status of the publication, indicating whether it is a refereed or nonrefereed publication. The few exceptions to this approach are those for which the ADS already contains a curated bibliography in its database, in which case the query just needed to retrieve those entries from this bibliography for the year range considered. Table 32.2 shows an overview of publication number for the start and end of the period, for a number of entities (with a wide range in size).

We will work with a dataset with maximum homogeneity by restricting both usage data and publication data to the main astronomy journals. This way we can be assured that all necessary metadata will be available for article selection.

32.3 Usage and Research Activity

One of the products of research activity is the generation of scholarly publications. It seems reasonable to assume that, as part of the preparation process, publications are read and may get added to the bibliographies of future publications. Since the ADS is the main literature discovery tool in astronomy, we assume that the publications read during this preparation process are found using the ADS. This model implies a number of questions:

1. Is there a correlation between the publications generated by an entity and the ADS usage by people associated with that entity?
2. Assuming that at least a fraction of downloaded papers will get cited in the publications generated, is there enough signal to detect that relationship?

In addition to the question of whether there is a correlation between downloads and number of publications, we also should look into the similarity of both sets. If our assumption is true that people download publications that serve as foundations for their publications, you would expect an overlap between the two sets. This leads to a third question:

3. How similar are the set of publications cited by the publications generated by an entity and the publications downloaded by people associated with that entity?

Before exploring these questions, we need to address the relationship between the actors in the usage and research activities. In the section on the data, we looked at usage data and publication data as separate quantities, but since the ADS lies at the center of both, we expect some common sense correlations. For the authors associated with a given entity, especially when they are the first author, we expect them to be among the frequent ADS users for that entity. This should be particularly true when we restrict ourselves to publications downloaded by the frequent ADS users associated with that entity. For example, given all the papers published in one of the main astronomy journals in 2005, where one of the authors is

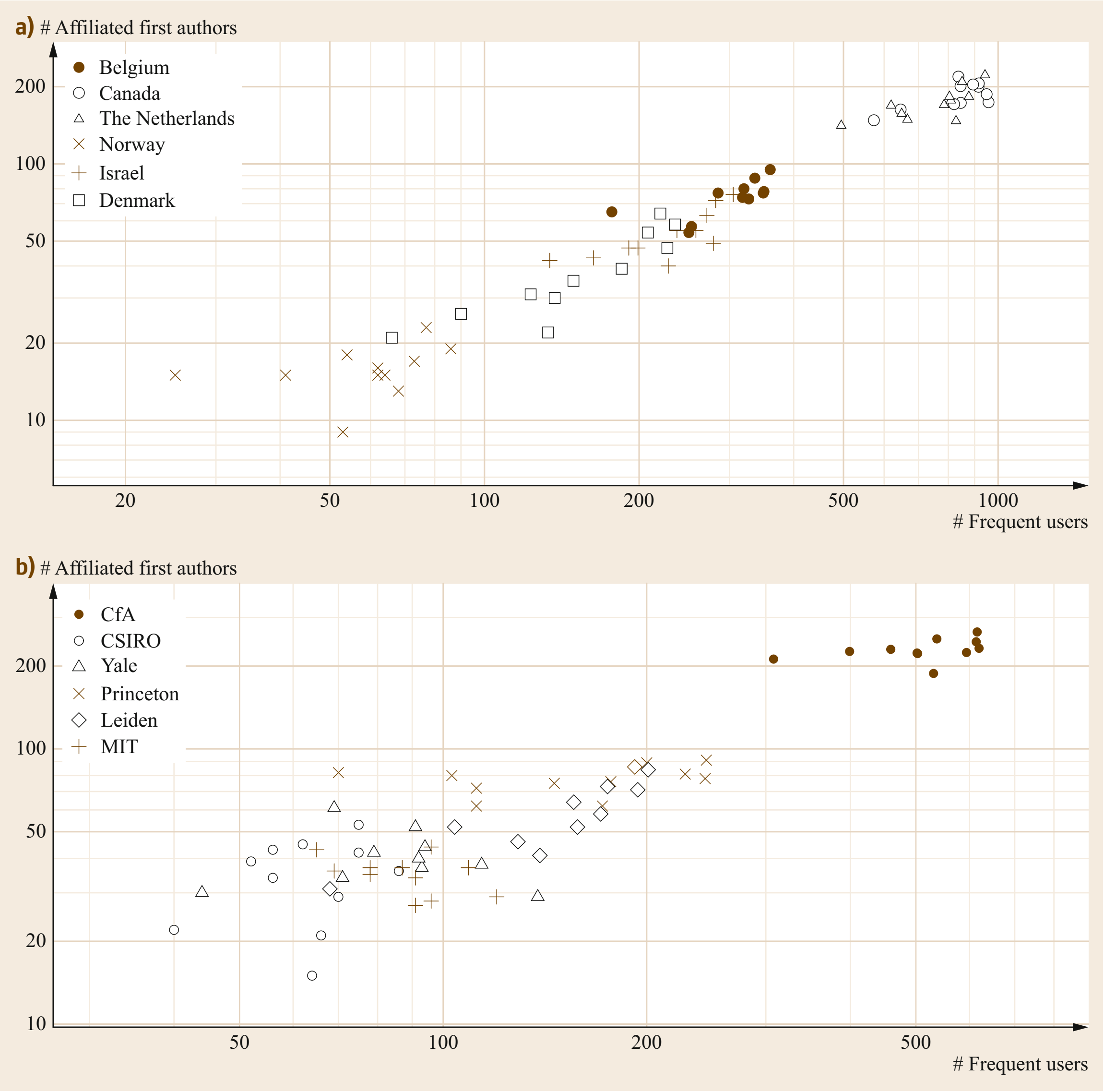


Fig. 32.4a,b Relationship between the number of frequent users and the number of affiliated first authors for a number of entities. Every data point corresponds with one year for the entity displayed. **(a)** Countries. **(b)** Institutes

affiliated with an institute in Belgium, what is the correlation between number of authors who appeared as first author in this set and the number of frequent users from Belgium in 2005? Figure 32.4 shows this relationship for a number of countries and institutes.

For countries, we also expect a significant correlation between the number of affiliated first authors and the number of national members of the IAU. This is shown in Fig. 32.5.

Figures 32.4 and 32.5 show that our data is consistent with common sense expectations: there is a significant overlap between the population downloading publications on a regular basis and publishing articles in main astronomy journals.

Next, we will explore the relationship between downloads and generated publications for a range of entities. As before, we will focus our analysis on frequent users and publications from the main astronomy journals. Figure 32.6 shows the numbers of downloads by frequent users of main astronomy papers recently published versus the number of main astronomy papers where one of the authors is affiliated with the entity under consideration.

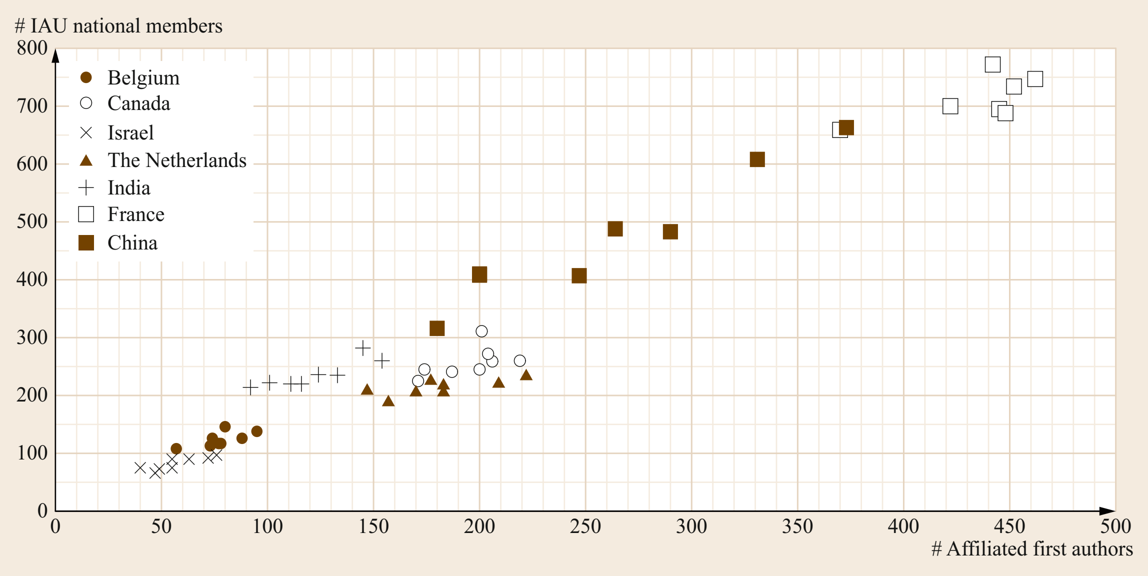


Fig. 32.5 Relationship between the number of affiliated first authors for a number of countries and the number of national members of the IAU. Every point represents a year in the period 2008–2015 (no IAU data was available prior to 2008)

The correlation between download and publication numbers has a scalar character. It does not provide any insight into the similarity of the downloaded and published articles. This is an aspect we can explore in a number of ways. Table 32.3 shows the quantities that will be used to explore the similarity between downloaded publications and those cited in a given year Y for a specific entity E.

Every publication has a publication year. Figure 32.7 shows a number of relationships based on publication year. One signal we have observed is, for a given year and entity, the set of downloaded publications in the main astronomy journals, by frequent users, with a publication year in the interval starting in 1980 and ending in the year under consideration. For this signal we derive the following two quantities. First, for each year in the range of publications years, the number of downloaded publications, normalized by the total number of publications in that set. The other quantity is the number of unique publications with a given publication year, normalized by the total number of publications in the main astronomy journals for that year. We derive the same quantities for another signal: all papers from the main astronomy journals and a publication year in the same range, cited by the publications in the bibliography for the entity in the year under consideration. Figure 32.7 shows the results for The Netherlands (E) in 2015 (Y).

Finally, we consider similarity on the most granular level: using individual publications as the data in the comparison. Assuming that researchers, affiliated with an entity, actually read the publications (at least to a nonnegligible degree) they cite in their scholarly papers, you would expect an overlap in those citations and the publications downloaded by frequent users from that same entity. The critical assumption here is that the ADS is used to get to the full text (either hosted locally on the ADS servers or external to the ADS). For both sets of publications, we have lists of identifiers, uniquely associated with publications. We will use these identifiers in our analysis. For a given entity in a given year, the main astronomy papers cited in the first-authored papers are compared with the downloaded main astronomy papers by frequent users, associated with the same entity. We will consider the similarity measured by the fraction of overlap between the two sets. Figure 32.8 shows results for a number of countries.

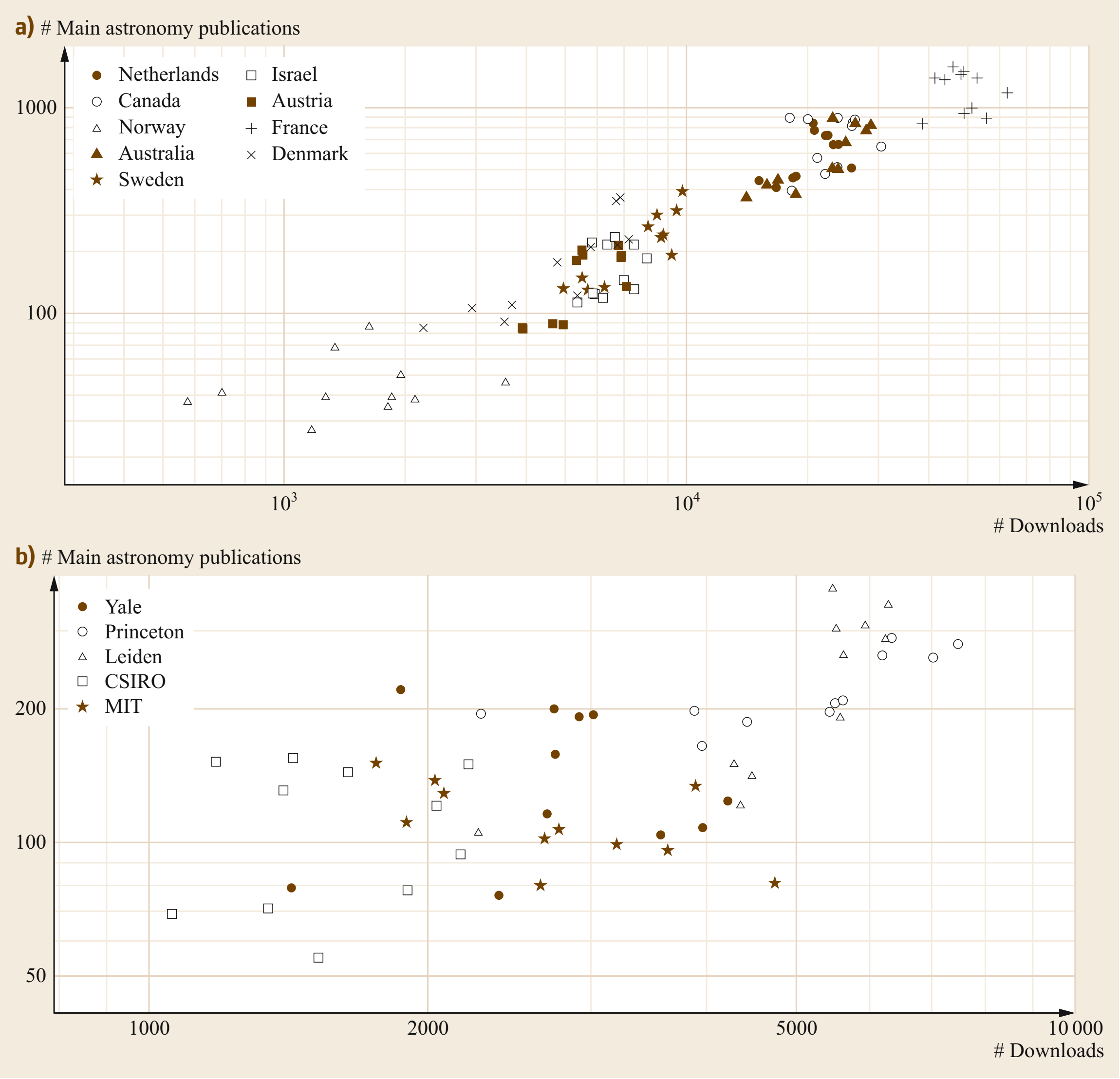


Fig. 32.6a,b Number of downloads by frequent users of main astronomy papers recently published versus the number of main astronomy papers where one of the authors is affiliated with the entity under consideration. Every point represents a year in the period 2005–2015. **(a)** Relationship for countries. **(b)** Relationship for institutes

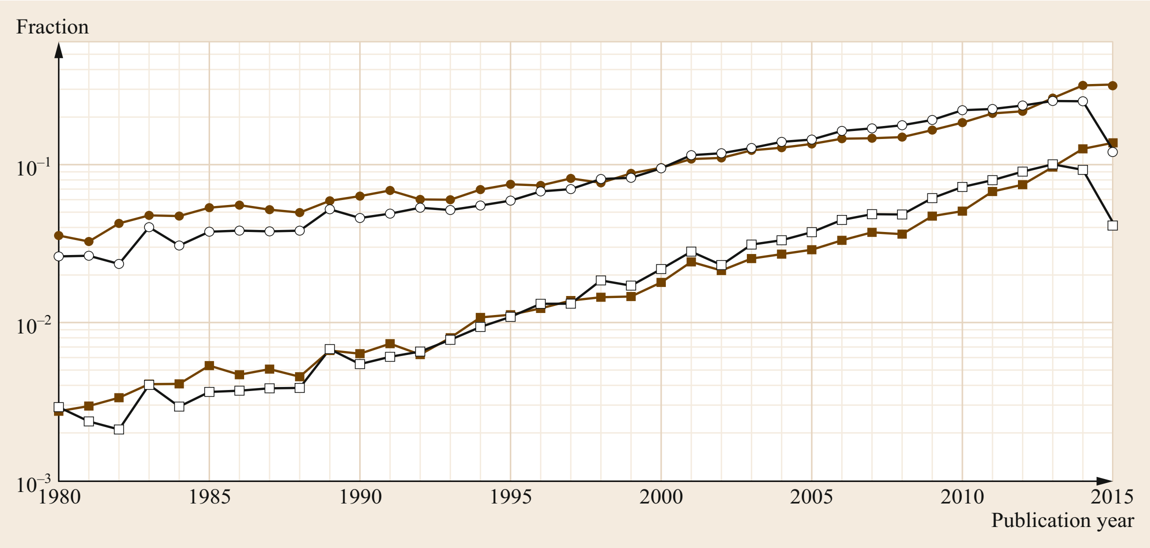


Fig. 32.7 For The Netherlands (entity) and the year 2015, this figure compares the distribution of publication years, in the range of 1980 through 2015, in the lists of downloaded main astronomy publications by frequent users and those cited by main astronomy publications in that year. The lines with *open* and *closed circles* show the fraction of unique publications in the citation (*open*) and download (*closed*) lists. The lines with *squares* show the normalized publication numbers in the citation (*open*) and download (*closed*) lists

Table 32.3 Definition of quantities used to explore the similarity between downloaded publications and those cited in a given year Y for a specific entity E

R^Y(E)	Publications in main astronomy journals downloaded by frequent users associated with E
P^Y(E)	Publications in main astronomy journals where the first author is affiliated with E
C^Y(E)	All publications in main astronomy journals cited by the publications in P^Y(E)

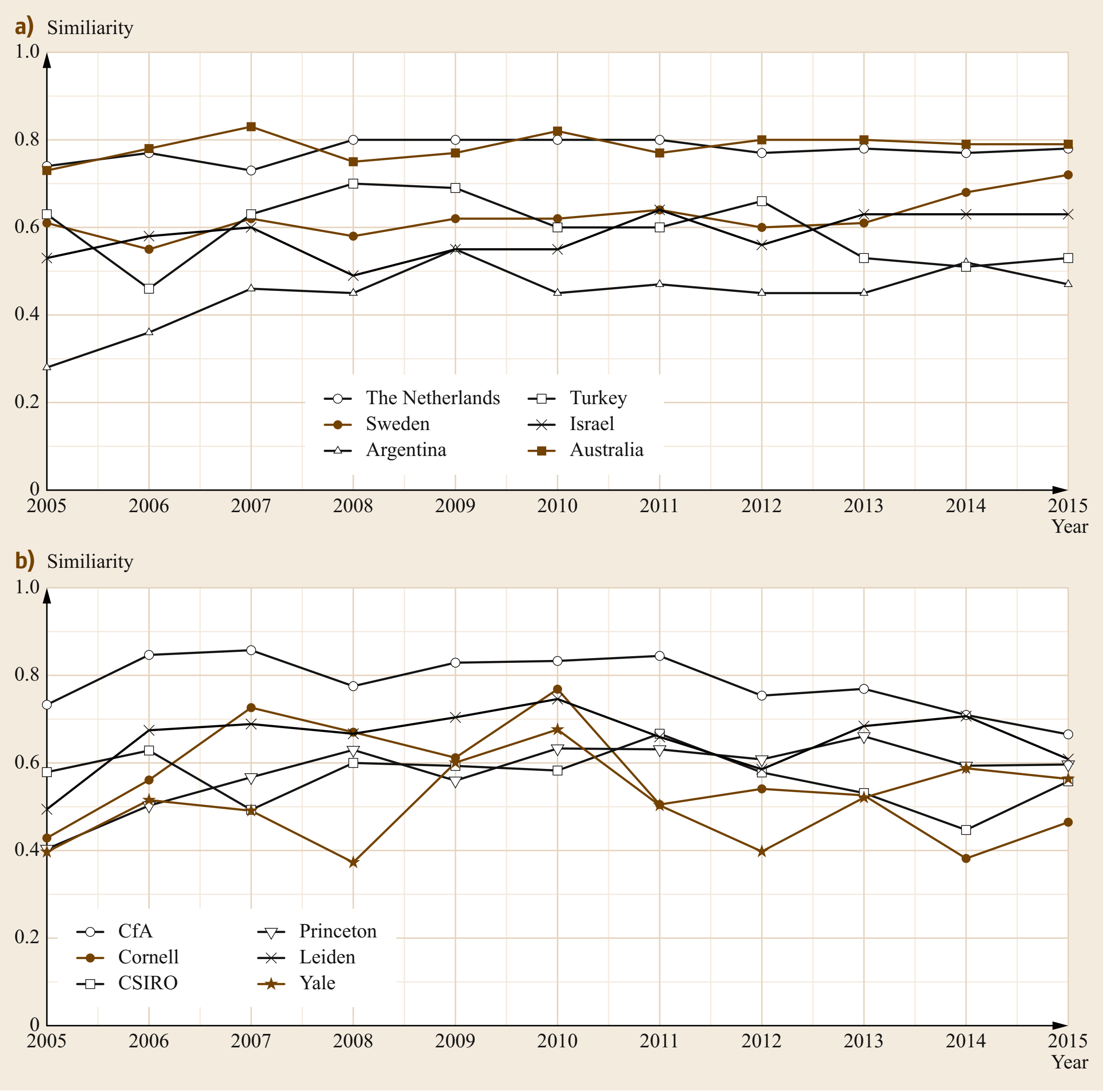


Fig. 32.8a,b The fraction of overlap between the sets of main astronomy papers cited in the first-authored papers are compared and the downloaded main astronomy papers by frequent users, associated with the same entity are shown. **(a)** Relationship for countries. **(b)** Relationship for institutes

32.4 Traditional Indicators

Bibliometrics is an example of a discipline that provides tools for quantifying research output and its impact. As mentioned before, we assume that research output, in the form of publications, can be regarded as a proxy for research activity. In this context, bibliometrics provides a suite of measures that can be seen as research metrics. These bibliometric measures are used to quantify a degree of research output. Traditionally, citations are used as fundamental building blocks for these measures. There are many citation-based research productivity measures, ranging from straightforward ones, like total citations, to complex indicators, like the Tori indicator [32.25]. The one thing they have in common is that, in fact, they are a measure for impact, rather than a direct measure for research activity. Of course, without research activity, there is nothing to cite, so in this sense they are an indirect measure for research activity. Even when e-prints from a repository like arXiv.org are used in conjunction with these measures, they do not provide the immediate

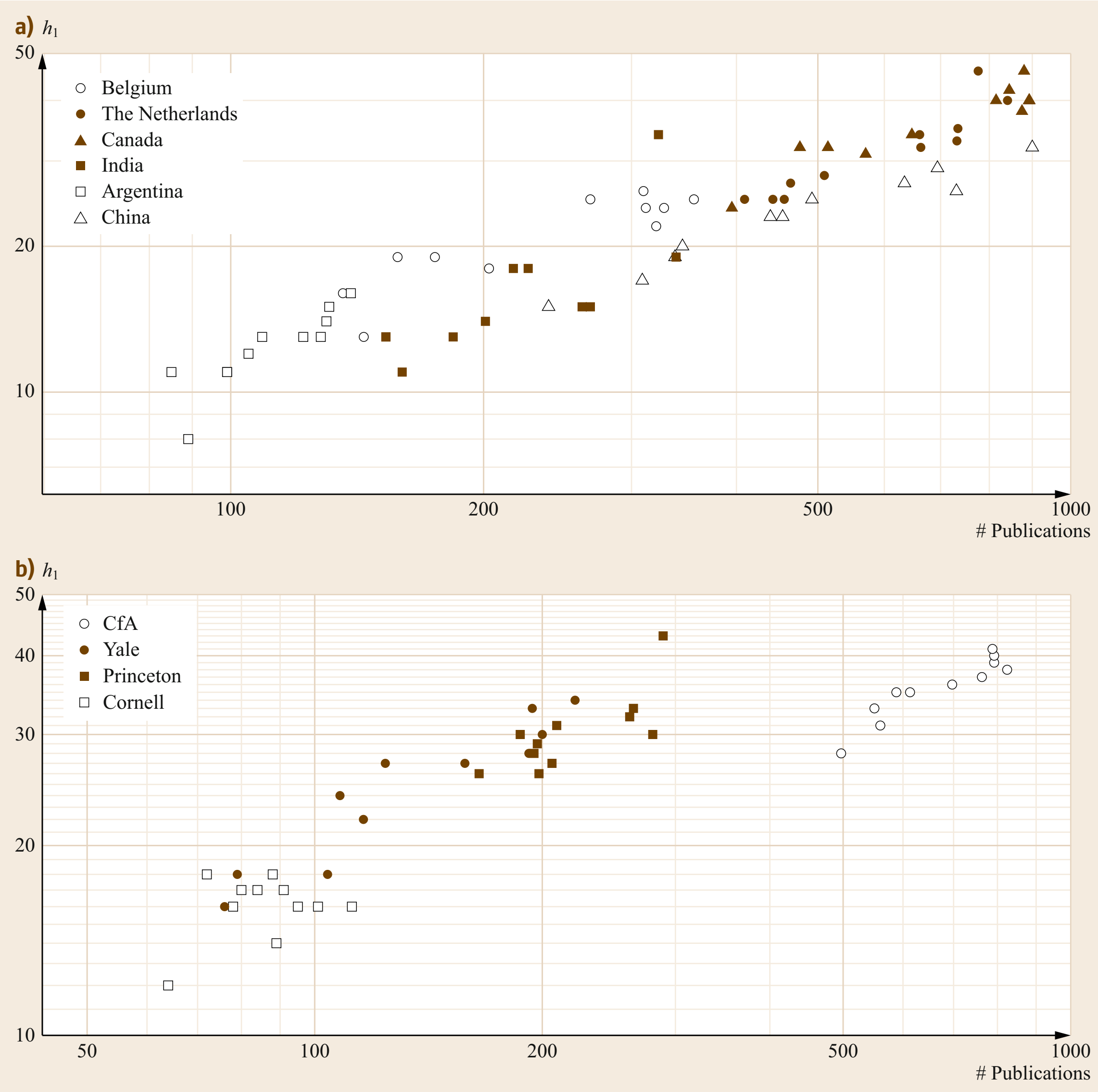


Fig. 32.9a,b The number of publications in the main astronomy journals associated with an entity compared to the h-index for these publication in the next year. **(a)** Countries. **(b)** Institutes

sense of research activity that usage-based measures provide. By comparing citation-based measures with those based on usage, you look at more than just research activity. This comparison also involves research quality.

With the correlation found between usage (in the form of downloads by frequent users) and publications, you would expect significant correlations with citation-based indicators. The act of publishing generates citations. In Fig. 32.9 we explore the relationship between the number of publications (in the main astronomy journals) generated for an entity in a year and the value of the h-index, for these publications, in the next year.

32.5 Discussion

Our selection of data for the analysis put forward in this publication has been guided by the following observation: while citation is a deliberate, public act, usage is private act, with a wide range of types. Citations are solely generated by authors, while usage is not solely the result of actions by authors. Taking publications as a measure for research activity, we necessarily need to focus on usage patterns associated with ADS users who are most likely to be authors. Because we have unique identifiers associated with users, we are able to do this. Figure 32.2 illustrates that, with a proper definition of usage frequency, a class of users can be delineated that strongly correlates with the size of professional researchers. With lower and upper boundaries set to 100 and 1000 downloads per year, over the period covered by this paper (2005–2015), the number of frequent users closely resembles the number of IAU members for the period with available data (2008–2015). For a number of entities, we determined the set of publications in the main astronomy journals where one of the authors has an affiliation associated with that entity. From these sets, we created subsets where the first author is affiliated with a particular entity. For each year in our period of analysis, we compared the number of first, affiliated authors with the number of frequent users for that entity. This comparison shows a very strong correlation for countries and a strong correlation for institutes, but definitely more spread. Examples of sources that contribute to spread are the fact that not all frequent users are authors and time lapses inherent to the act of publishing (particularly relevant for publications early and late in each year). In countries with multiple institutes, especially those with higher numbers of frequent users and authors, spread is less because of an averaging effect. In the case of countries, Fig. 32.5 compares the number of affiliated first authors with the number of national IAU members, for the period 2008–2015. Affiliated first authors can be visiting scholars from another country or simply not be a member of the IAU. Also, not all national IAU members are active authors. Nevertheless, Fig. 32.5 shows a strong correlation between both numbers. Taken together, Figs. 32.2, 32.4 and 32.5 make a strong case for being able to trust that the class of frequent users, derived from the ADS usage logs, sufficiently represents those scholars that generate research activity in the form of publications, for a given entity.

In the next stage, we compare the signals, at various levels of granularity, generated by the two populations we have defined: authors affiliated with an entity and frequent users associated with that entity. The author-centric data consists of publications generated in a particular year (and the publications cited in their respective bibliographies) and the data for the frequent users consists of lists of publications downloaded by them in a particular year. In both cases, publications are identified by a unique entity. At the least granular level, we compare just numbers, specifically the number of publications from the main astronomy journals, downloaded by a frequent user associated with an entity, and the number of main astronomy publications where one of the authors is affiliated with that entity. Figure 32.6 shows that the correlation is the strongest for countries. As noted before, there are the effects of missed data at the start and end of each year, but the effect of this gets dampened on the scale of a country. The smaller amounts of data on the level of institutes results in a larger amount of scatter. At best, the results in Fig. 32.6 show that there is a relationship in the mean. By comparing scalar quantities we have removed any information on what actually is downloaded and published.

Ideally, authors should read every publication that they intend to cite in the paper they are writing. However, this is unlikely to be true [32.26]. The question is whether this is true enough. Assuming that the populations of frequent users sufficiently represent authors, we still expect a lot of scatter on the detailed level of actual downloaded publications. Publications are downloaded that for various reasons do not get cited or get cited in a paper which appears in the next year. Figure 32.7 explores relationships at a more granular level, but not yet at the most granular level of using individual publications in comparisons. Figure 32.7 was constructed by using one particular aspect of article metadata: the publication year. With the publication year we examine the similarity in obsolescence functions for cited and downloaded publications. We explored these quantities for one particular entity and year, but the results were found to hold true for a wide range of either. The similarity of the obsolescence functions indicates that the correlation found earlier, in the least granular level, is still present at the more granular level of using the publication years of individual articles.

Before going to the level of individual publications, there is at least one additional approach for comparing the results for both populations. This approach would give an estimate for the similarity in subject matter between the published publications and those downloaded by frequent users, for example by using a clustering algorithm (like k-means or Louvain clustering based on keywords). We will not consider this approach in this publication, but it may be a subject for a future publication.

Finally, we explore the similarity between downloaded publications and the cited papers, for a range of entities. We selected the most straightforward comparison: calculating the fraction of overlap between both sets. Figure 32.8 shows the results for a number of entities. Would these fractions have been significantly smaller if the downloads had been a random selection of publications from the main astronomy journals? Figure 32.10 shows results for a number of entities. Every *random* value is the average of ten samples.

For both countries and institutes we find moderate, yet significant similarities between main astronomy publications, downloaded by frequent users and the main astronomy publications cited in the publications for a given entity and year. We wouldn't expect this similarity to be more than modest because of the reasons we mentioned earlier.

Is there a relation with traditional bibliometric measures? It is difficult to use traditional bibliometric measures to quantify research activity, in particular if you are interested in a measure that is very close, temporally, to when the research is actually happening. For any measure based on citations, the articles will have to have been published and available long enough to accumulate a reasonable amount of citations. By considering citations as a measure, you are immediately attaching a quality assessment to your analysis. A usage-based measure, like the READ10 indicator [32.22], based on current reads of publications, is an alternative, but it is still an impact measure. This does not mean that there is nothing to compare. Given the fact that most of these traditional measures are citation-based, any relationship is indirect and circumstantial at best. We have shown that download activity correlates with publication numbers. Publications result in citations. Taking

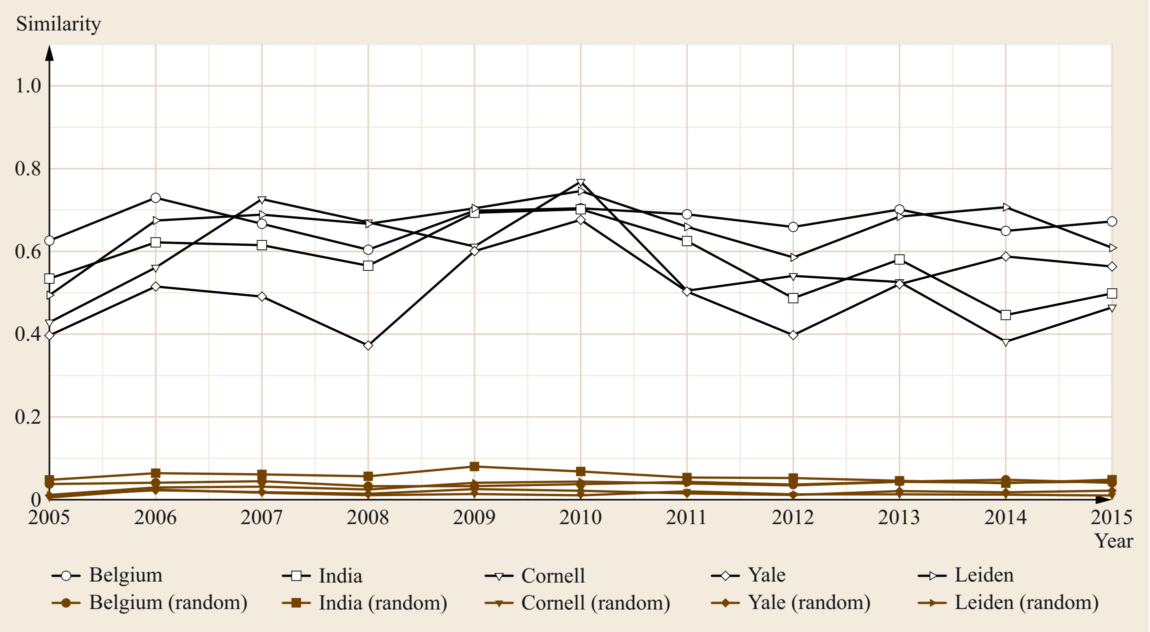


Fig. 32.10 The fraction of overlap between the sets of main astronomy papers cited in the first-authored papers are compared and the downloaded main astronomy papers by frequent users, associated with the same entity are shown. Lines with *open symbols* show the actual data and lines with corresponding *solid symbols* show the results if the downloads had been random

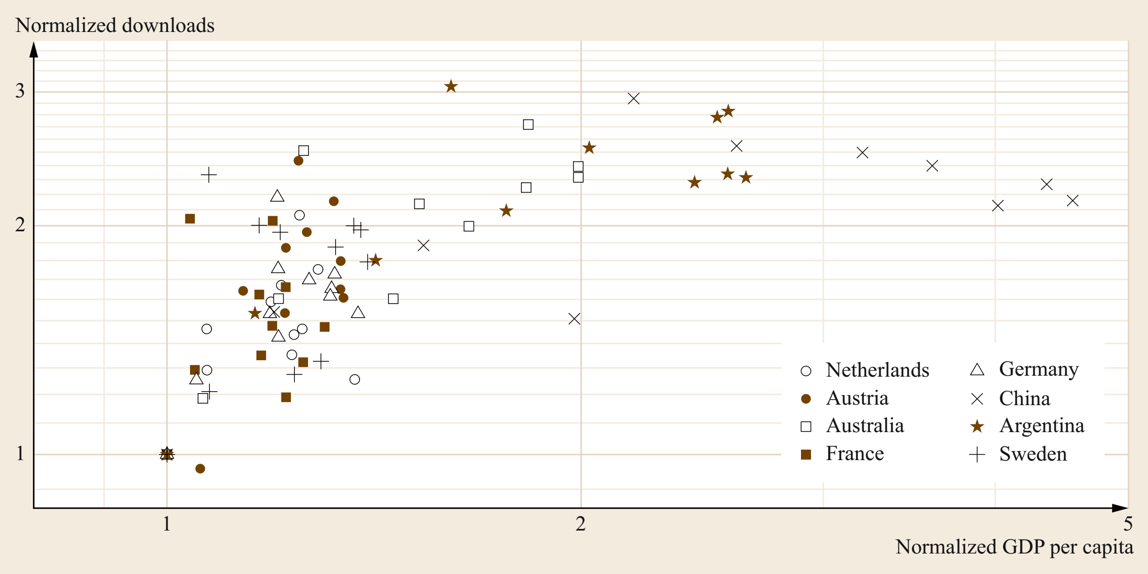


Fig. 32.11 GDP per capita (current USD) versus number of downloads by frequent users. Both quantities have been normalized by their values in 2005. Every point corresponds to a year in the range 2005–2015

into account that there is a lag for citations to accumulate, there may be a correlation between current downloads and a citation-based indicator *later* (say, the following year). The results in Fig. 32.9 show this to be a correct assumption. *Kurtz* et al. [32.22] showed the relation between downloads of an author's work and citations to that work, both on an individual author basis, and summed over research institutes.

Eichhorn et al. [32.21] first showed the relation between the number of times individuals in a country download an astronomy article and the GDP of that country, and in an earlier publication we showed a relationship between general usage and the GDP per capita for a range of geographic regions [32.24]. We found that growth in GDP per capita in general translates to an increase in ADS usage. Is there a similar trend for just downloads by frequent users? Figure 32.11 indicates that, with a few exceptions, an increase in GDP per capita results in an increase of downloads by frequent users, implying increased research activity. An exception is China, where it seems that research activity does not keep up with the explosive economic growth.

Kurtz et al. [32.23] showed that the number of downloads in a country is proportional to that country's GDP squared divided by it population. In that publication, they also showed that that the number of downloads by researchers of astronomy articles better predicts the mean of the number of articles published and the citations of those articles—all measured as fractions of the world total.

32.6 Concluding Remarks

When a researcher cites a publication in an article, it is a public, deliberate act. The only room for interpretation is the sentiment of the citation. From a process point of view, because of this deliberate nature, there is no noise component in the citation signal. This cannot be said for the process of usage. For any entity where some level of scholarly research is performed, literature discovery is an essential ingredient in the research lifecycle. However, this is a process with many components, depending on the goal of the literature search. All of this activity, for all entities on all levels, comes together for a service provider like the ADS. Somewhere, buried in the millions of yearly interactions with the service, are the signals that represent the act of gathering the necessary literature for producing scholarly articles. Although *research activity* consists of many components that do not involve literature search and writing, we feel that the act of generating scholarly articles is a measure that is a good proxy for research activity. Key to being able to find these signals is the ability to associate sessions with individuals. It is not sufficient to use IP addresses to identify sessions. For one, researchers are mobile and some institutes proxy their requests through one single host. We noted the presence of library computers in Fig. 32.2 and how these result in usage of multiple users end up being indistinguishable. For usage information to have any use, either for bibliometric analysis or to enhance the discovery experience of a service, a lot of meaningless systematic (e. g., robots) and random (e. g., incidental users entering through Google) signals need to be removed. We showed the dramatic difference in orders of magnitude between incidental use and the use by people who use the ADS professionally. When thinking about a very specific class of researchers, namely authors, it is important that all authors use the ADS, but that not all users (even frequent users) are authors. Going back to our original question of quantifying research activity, this emphasizes the need for being able to properly identify these frequent users. In this study we showed that for the ADS we can identify a class of frequent users and that there is convincing evidence that these users represent the population of active researchers (and even authors) in astronomy. Having identified these frequent users for various entities and having constructed bibliographies for these entities for the time period of 2005–2015, we have shown that there is a significant correlation between the number of frequent users and first, affiliated authors for a range of entities. For countries we also showed a correlation between the number of first, affiliated authors and the number of national IAU members. Based on this evidence we argue that the class of frequent users represents the authors making use of the ADS while in the process of writing scholarly articles. For the main astronomy journals, the download activity by these frequent users correlates with publications on multiple levels of granularity. This correlation is stronger for countries than for institutes. Our conclusion is that download activity for main astronomy journals represents research activity. Even though we did not show this explicitly, we feel that this observation can be extended to refereed literature in general. The research cycle is a process common to many disciplines, and producing scholarly publications is always part of this cycle. Since all authors in astronomy are also users of the ADS and always use the ADS in their research, we feel that our results, presented in this paper, are of a more general nature and not just an indicator of trends in astronomy.

Is it meaningful to consider rankings of entities based on usage-based indicators? Making meaningful comparisons using citation-based indicators is already a complicated issue [32.18], so doing this using data that is intrinsically more noisy is going to be very hard indeed and probably even meaningless in a practical sense.

Acknowledgments. This work has been supported by the NASA Astrophysics Data System project, funded by NASA grant NNX16AC86A.

References

32.1 National Science Board: Science & Engineering Indicators, https://www.nsf.gov/statistics/2016/nsb20161/#/report (2016)
32.2 Times Higher Education Ranking: https://www.timeshighereducation.com
32.3 Academic Ranking of World Universities (ARWU) http://www.shanghairanking.com
32.4 E. Garfield: The history and meaning of the journal impact factor, J. Am. Med. Assoc. **295**, 90–93 (2006)
32.5 Eigenfactor: http://www.eigenfactor.org
32.6 C.T. Bergstrom, J.D. West, M.A. Wiseman: The Eigenfactor™ metrics, J. Neurosci. **28**, 11433–11434 (2008)
32.7 H. Moed: Measuring contextual citation impact of scientific journals, J. Informetr. **4**, 265–277 (2010)
32.8 COUNTER: https://www.projectcounter.org
32.9 M.J. Kurtz, G. Eichhorn: The historical literature of astronomy, via ADS, ASP Conf. Ser. **153**, 293–296 (1998)
32.10 M.J. Kurtz, J. Bollen: Usage bibliometrics, Annu. Rev. Inf. Sci. Technol. **44**, 3–64 (2010)
32.11 D. Nicholas, P. Huntington, T. Dobrowolski, I. Rowlands, M. Jamali, P. Polydoratou: Revisiting "obsolescence" and journal article "decay" through usage data: An analysis of digital journal use by year of publication, Inf. Process. Manag. **41**, 1441–1461 (2005)
32.12 D. Nicholas, D. Clark, H.R. Jamali, A. Watkinson: Log usage analysis: What it discloses about use, information seeking and trustworthiness, Int. J. Knowl. Content Dev. Technol. **4**, 23–37 (2014)
32.13 M. Thelwall, K. Kousha: ResearchGate articles: Age, discipline, audience size, and impact, J. Assoc. Inf. Sci. Technol. **68**, 468–479 (2017)
32.14 C. Tenopir, L. Christian, R. Anderson, L. Estelle, S. Allard, D. Nicholas: Beyond the download: Issues in developing a secondary usage calculator, Qual. Quant. Methods Libr. **5**, 365–377 (2016)
32.15 W. Glänzel, J. Gorraiz: Usage metrics versus altmetrics: Confusing terminology?, Scientometrics **102**, 2161 (2005)
32.16 J. Gorraiz, C. Gumpenberger, C. Schlögl: Usage versus citation behaviours in four subject areas, Scientometrics **101**, 1077 (2014)
32.17 H.F. Moed, G. Halevi: On full text download and citation distributions in scientific-scholarly journals, J. Assoc. Inf. Sci. Technol. **67**, 412–431 (2016)
32.18 M.J. Kurtz, E.A. Henneken: Measuring Metrics – A forty year longitudinal cross-validation of cita-

tions, downloads, and peer review in astrophysics, J. Assoc. Inf. Sci. Technol. **68**, 695–708 (2017)
32.19 J. Bollen, H. van de Sompel: Usage impact factor: The effects of sample characteristics on usage-based impact metrics, J. Am. Soc. Inf. Sci. Technol. **59**, 136–149 (2008)
32.20 C.J. Lintott, K. Schawinski, A. Slosar, K. Land, S. Bamford, D. Thomas, M.J. Raddick, R.C. Nichol, A. Szalay, D. Andreescu, P. Murray, J. Vandenberg: Galaxy Zoo: Morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey, Mon. Not. R Astron. Soc. **389**, 1179–1189 (2008)
32.21 G. Eichhorn, M.J. Kurtz, A. Accomazzi, C.S. Grant, S.S. Murray: The NASA Astrophysics Data System: The search engine and its user interface, Astron. Astrophys. Suppl. **143**, 61–83 (2000)
32.22 M.J. Kurtz, G. Eichhorn, A. Accomazzi, C. Grant, M. Demleitner, S.S. Murray, N. Martimbeau, B. Elwell: The bibliometric properties of article readership information, J. Am. Soc. Inf. Sci. Technol. **56**, 111–128 (2005)
32.23 M.J. Kurtz, G. Eichhorn, A. Accomazzi, C. Grant, M. Demleitner, S.S. Murray: Worldwide use and impact of the NASA Astrophysics Data System digital library, J. Am. Soc. Inf. Sci. Technol. **56**, 36–45 (2005)
32.24 E.A. Henneken, M.J. Kurtz, A. Accomazzi, C.S. Grant, D. Thompson, E. Bohlen, S.S. Murray: Use of astronomical literature – A report on usage patterns, J. Informetr. **1**, 1–8 (2009)
32.25 A. Pepe, M.J. Kurtz: A measure of total research impact independent of time and discipline, PLoS ONE **7**, E46428 (2012)
32.26 M. Simkin, V. Roychowdhury: Do you sincerely want to be cited? Or: read before you cite, Significance **3**, 179–181 (2006)

Edwin Henneken

High Energy Physics
Harvard-Smithsonian Center for Astrophysics
Cambridge, MA, USA
ehenneken@cfa.harvard.edu

Edwin Henneken studied Astrophysics at Leiden University and Micro-Meteorology at the Vrije Universiteit Amsterdam. After working in the Internet industry, he joined the Astrophysics Data System team as an IT Specialist at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts. He uses the ADS for bibliometrics research, literature recommendation, and exposure of products such as data and software.

Michael Kurtz

Optical and Infrared Astronomy
Harvard-Smithsonian Center for Astrophysics
Cambridge, MA, USA
mkurtz@cfa.harvard.edu

Michael Kurtz is an astronomer and computer scientist at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts. Kurtz is the author or co-author of over 300 technical articles and abstracts on subjects ranging from cosmology and extra-galactic astronomy to data reduction and archiving techniques to information systems and text retrieval algorithms. He is the project scientist for the Smithsonian/NASA Astrophysics Data System.

33. Online Indicators for Non-Standard Academic Outputs

Mike Thelwall

This chapter reviews webometric, altmetric, and other online indicators for the impact of nonstandard academic outputs, such as software, data, presentations, images, videos, blogs, and grey literature. Although the main outputs of academics are journal articles in science and the social sciences, and monographs, chapters, or edited books to some extent in the arts and humanities, many scholars also produce other primary research outputs. For nonstandard outputs, it is important to provide evidence to justify a claim for a type of impact and online indicators may help with this. Using the web, academics may obtain data to present as evidence for a specific impact claim. The research reviewed in this chapter describes the types of evidence that can be gathered, the nature of the claims that can be made, and methods to collect and process the raw data. The chapter concludes by discussing the limitations of online data and summarizing recommendations for interpreting impact evidence.

33.1 Non-Standard Academic Outputs

Although most scientometric evaluations of researchers are restricted to journal articles, there are many other valid academic outputs. In the humanities, monographs are more important than journal articles and edited books are also valued [33.1, 2]. Similarly, conferences can be more important than journals in engineering-related fields, and performances, exhibitions, and individual works of art are key academic outputs in the arts. Thus, in some fields, journal articles are secondary or irrelevant. Even in disciplines where the traditional journal-based publishing model dominates, individual scholars may legitimately focus on other outputs and activities that make important contributions to the wider goals of science, such as patents, educational resources, software, data, and websites as well as public outreach blogs, videos, and presentations. It is important that these are not ignored or undervalued in research evaluations because this would narrow the focus of science and isolate it from its wider societal goals. It would also make science less efficient if communal activities like data and software sharing were inadvertently discouraged.

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_33

Given that it is essential to recognize the value of nonstandard academic outputs, quantitative indicators are needed to support this. In practice, nonstandard outputs are probably evaluated informally without the aid of systematic data in appointment, tenure, and promotion decisions. If an academic's blog or software is well-known in their field then this might be enough to ensure that this contribution is valued when she applies for promotion. Similarly, an appointment committee may notice a list of media engagements on an applicant's curriculum vitae (CV) and informally estimate the value of this contribution by the length of the list and the prestige of the sources mentioned. Nevertheless, many critical decisions are made by senior committees that are not field experts and might find it difficult to evaluate nonstandard contributions to research. Moreover, most academics do not produce high-profile work, making the exact value of their outputs difficult to determine, even for subject experts. Thus, there is a need to obtain quantitative data to help the evaluation of nonstandard outputs.

One former EU (European Union) project took a researcher-centered approach to evaluations and argued that scholars should produce a portfolio of evidence of their outputs, achievements, and capabilities [33.3, 4]. Within this portfolio, individuals could list their outputs and any available quantitative impact evidence. The heart of their claim for value would be a narrative statement that refers to this evidence. This sets the quantitative evidence of a researcher's achievements in the context of the value that they believe that it has. One researcher might present 1000 website visits as evidence of substantial success for a blog aimed at field experts, whereas another might claim 100 000 visits as similarly powerful evidence of success for a blog aimed at science education for school pupils. The ACUMEN portfolio requirement for the researcher to marshal their evidence and make a claim for the value of their work formalizes the more standard process of job applicants submitting a cover letter and CV. The portfolio approach also foregrounds the importance of data to support claims.

Nonstandard outputs are prominently evaluated in the UK research excellence framework (REF) national research evaluations [33.5, 6]. As part of the 2014 evaluation, UK academics selected four outputs to represent their best work 2008–2013. Whilst most chose journal articles, monographs were common in the humanities and a range of different artistic outputs in arts subjects. In both cases the outputs were evaluated by subject experts' judgements of their significance, rigor and impact, without the aid of metrics. In contrast, REF2014 also required submitting departments to construct impact case studies, which were narratives describing how some of the department's previous research had made an impact outside of academia [33.7]. These narratives were expected to provide evidence of impact and this could take the form of qualitative support or quantitative data from any source. The impact case studies provide many examples of how ad hoc quantitative data can be used to support impact claims. The following extracts illustrate this:

- Claim: "Ulinka Rublack's research focuses on the history of Renaissance dress. Her work has enhanced public awareness that social groups beyond courtly elites created fashion in the past. It led to a re-creation of one of the most significant outfits recorded in the wardrobe of a sixteenth-century accountant." [33.8] Evidence example: "Cambridge University Website article on Rublack's research and the reconstruction, launched 1 May 2013, shared 2051 times, tweeted 253 times, taken up as top item on website http://medievalists.net" [33.8]
- "[Breeze's] research into rich donors in the UK has impacted on the policies of both governmental and non-governmental bodies." [33.9] Evidence example: "Breeze has 2250 followers on her @UKCPhilanthropy Twitter account, and has tweeted 1306 times (as at 24/10/13)." [33.9]
- "The 10 min Puzzle podcast series seeks to engage lay audiences with some of the central puzzles driving contemporary research in analytic philosophy. As of September 19th, 2013, there had been over 63 000 downloads." [33.10]
- "Visual analytics is a powerful method for understanding large and complex datasets that makes information accessible to non-statistically trained users. The Non-linearity and Complexity Research Group (NCRG) developed several fundamental algorithms and brought them to users by developing interactive software tools (e. g., Netlab pattern analysis toolbox in 2002 (more than 40 000 downloads))." [33.11]
- "Dr Katharina Hall's blog Mrs Peabody Investigates (http://mrspeabodyinvestigates.wordpress.com/; henceforth MPI) has been fostering public debate on German, European and international crime fiction since January 2011. Beneficiaries include readers, authors, translators, publishers, critics and bloggers in 130 countries. With over 220 000 hits and 2500 comments, MPI has been featured on BBC Radio 4 and is linked to by BBC Online, crime blogs, and publisher/author websites (C10). Providing a distinctive service of academically-informed reviews of high-quality crime fiction, MPI

is regarded in the industry as *a ground-breaking blog that is transforming readers' understanding and appreciation of international crime* (The Times crime-fiction critic)." [33.12]

- "In collaboration with film-maker Brady Haran we have developed the YouTube channel *60 Symbols* to present topics related to research in physics to the wider public. Since the 2009 launch of *60 Symbols* we have posted 212 videos, which have amassed 21.2 M views, over 200 k comments, over 266 k subscribers and a content approval rating of 99.4%, placing *60 Symbols* in the top 0.01% of all YouTube channels." [33.13]

Of course, there are many academic contributions that cannot be easily quantified. For example, the REF impact case study, *Preventing disease through promotion of handwashing with soap* (HWWS), makes the hugely impressive claim [33.14]:

> [Val] Curtis has spearheaded an effective alliance of industry with organisations like the World Bank, USAID and UNICEF to promote good hygiene. This means that millions of people around the world have now been exposed to HWWS promotion programmes. In the last decade, diarrhoea deaths in under 5 s have steadily fallen from 1.2 to 0.85 m per year, and while some of the credit must go to economic development and improved clinical treatment, some is undoubtedly due to the promotion of better hand hygiene based on Curtis' research.

The concrete claim *has spearheaded* is clearly one that is difficult to provide quantitative data to support and the impact data provided covers the work of the *alliance* rather than the individual researcher. In this case, and probably most others, it is not possible to quantify individual contributions by academics.

The ACUMEN portfolio and REF case study examples to some extent circumvent a major difficulty in using quantitative data for nonstandard research outputs. Because such outputs are nonstandard, with varied purposes as well as different forms and intended audiences, it is difficult to benchmark the results or to normalize them in any way [33.15]. In consequence, it can be unclear whether an indicator value represents high or low impact. There is no equivalent to the field-normalized citation count indicators for journal articles, where values above 1 indicate impact above the world average for the field and year examined [33.16]. There is not a simple solution to this problem for most of the output types discussed here. A fundamental issue in this context is that whilst journal articles address very broadly similar-sized audiences (other scholars in the same field), nonstandard outputs could be intended to reach a small audience, such as other specialist scholars, or a very large audience, such as members of the public interested in science. It would be unfair to directly compare indicators for outputs designed for such different-sized audiences.

The lack of field normalization for typical nonstandard outputs makes it more natural to evaluate them individually rather than through a group average. Effective normalization is needed for group averages to make sense. Thus, for example, it would not be helpful to report the average visitor count for all blogs produced by a university, whereas the visitor counts of each individual blog alongside information about its purpose and intended target audience could be useful.

This chapter discusses webometric, altmetric, and other methods to evaluate a range of nonstandard academic outputs. Webometric methods refer to techniques to extract impact evidence from the web in general [33.17]. In contrast, altmetric methods are restricted to data from social websites [33.18, 19], although the term seems to now be used to encompass all online indicators. Social web data is typically easier to obtain and more plentiful but is not always better than webometric data. This chapter extends a previous literature review of this topic [33.20] with updated and wider coverage. It also includes evaluations of websites, as researched within the field of webometrics. It does not cover all academics' activities. It excludes most awareness-raising activities, such as tweeting, maintaining a personal home page, or engaging on ResearchGate or Academia (http://www.academia.edu). Whilst these are all important activities, they are rarely the primary outputs of scholars.

33.2 Core Concepts

All online indicators need to be collected and evaluated before they can be used with confidence. These are discussed here to support future sections on specific indicators.

33.2.1 Indicator Creation

An impact indicator for a nonstandard output is a number that is expected to associate with the impact of the output. For example, the viewer count of a scientific YouTube video might be used as an impact indicator because, whatever type of impact the video has, the more people watch it, the more impact of that type it is likely to generate.

Some nonstandard outputs have natural and easily accessible indicators, such as the view count visible on YouTube video pages. In this case, if an indicator is needed for a single output then it can be looked up in the hosting website. If indicator values are needed for large collections of outputs, including for testing, evaluation or benchmarking purposes, then an automatic process is needed to gather the numbers. Data collection for informetric purposes is discussed in another chapter but the basics are also covered here.

For websites like YouTube with an application programming interface (API), it can be possible to automatically gather indicator values through a computer program. An API is a set of facilities within a website that allows computer programs to automatically access some or all the website content. Using an API, a researcher could feed a program with a list of nonstandard outputs and then it would download their scores. At the time of writing, the free software Webometric Analyst (http://lexiurl.wlv.ac.uk) could use the YouTube API to gather usage data for YouTube videos, the Twitter API to gather retweet and follower count data for individual users, the Mendeley API to count the registered Mendeley readers of any type of document, and the Google Books API to count citations from books. The Altmetric API (https://api.altmetric.com/) can also be used to download impact information from a variety of sources for individual outputs. It is mainly restricted to items with a digital object identifier (DOI), which most nonstandard outputs do not have. For websites with an API that is not supported by any currently existing software for academic purposes, a user would need to create a new program for their task. This is a straightforward task for an experienced programmer.

Most websites do not have an API, making it more difficult to extract data automatically. Examples at the time of writing include ResearchGate and Academia.edu. For sites without an API the main alternative strategy is to use a web crawler and/or page scraper software to automatically extract relevant data. A web crawler is a computer program that can be fed with one or more URLs (uniform resource locators) and then downloads them, perhaps also recursively following the hyperlinks in the downloaded pages (typically to other pages in the same website). After downloading the pages, the indicator data can be automatically extracted from the pages using a custom-written program, called a page scraper, which can either be part of the web crawler or a separate utility. The free web crawler SocSciBot (http://socscibot.wlv.ac.uk) is an example of a general-purpose crawler that can download either entire websites or collections of pages within a website. It is paired with Webometric Analyst, which incorporates page scrapers for many different websites. At the time of writing, the combination of these two programs (i. e., first downloading with SocSciBot, then page scraping with Webometric Analyst) could be used to systematically extract indicator values from ResearchGate, Academia.edu, SlideShare, DataDryad.org, Google Code, and FigShare (see the *Webometric Analyst Services* menu). When using a web crawler, it is important not to overload the targeted website by crawling it too quickly or by ignoring a request not to crawl it [33.21]. Damaging an academic website in this way would be a serious problem. Ethical behavior of this nature is designed into SocSciBot and most web crawlers to ensure that no damage is inadvertently done.

Many repositories and individual outputs have no accessible indicators. For example, the mainly physics repository arXiv does not report download or view counts. Similarly, if researchers post a report to their own institutional website then they may not have access to any usage statistics for it. For these outputs, it may still be possible to gather online citation data by counting how often they have been mentioned on the web. If some uses of an artifact mention it online, counting web citations (i. e., the number of web pages mentioning it) is a reasonable way to generate a simple impact indicator. For individual outputs, the easiest way to achieve this is to Google their titles and then manually check the results for accuracy to get an estimate of the number of online mentions. This underestimates the number of online citations because commercial search engines do not reach or report the entire web for various reasons [33.5, 22–26] and have internationally biased coverage [33.27] but this underestimate may still give an idea of the order of magnitude of the number of results. It also allows comparisons between different outputs or sets of outputs.

It is possible to automatically gather data from search engines on the number of web citations for scholarly outputs using the Bing search API. Thus, to count the number of web citations for each one of many outputs (e. g., grey literature, presentations), appropriate Bing queries can be automatically submitted. The program Webometric Analyst incorporates code to interact with the Bing API. This includes procedures to create appropriate web citation queries for sets of documents as well as to submit them to Bing and summarize the results [33.28]. The initial step, converting the output information into Bing queries, typically uses the first words of the title as a phrase search and then adds the first author last name and the publication year. This is based upon the assumption that when an output is cited in any web page, its title, publication year, and first main author are likely to be mentioned. This is clearly an oversimplification because, for instance, news stories rarely give much information about research that they mention. Nevertheless, it is a simple way to construct searches. Whilst the Google custom search is not designed to provide a general search service (http://developers.google.com/custom-search/), it is possible to adapt it for this and so Bing is not the only choice for automatic queries.

For citation indicators (e. g., citations to grey literature or theses), Scopus or the Web of Science (WoS) may give citation counts but online alternatives, such as Mendeley [33.29], Google Scholar [33.30], Microsoft Academic [33.31, 32], or ResearchGate [33.33] may give earlier evidence of impact.

In summary, for individual outputs, impact evidence may sometimes be obtained from the download data or other usage statistics in the hosting website and, if this is not available, then search engine queries could be used to obtain (an underestimate of) the number of pages mentioning the output. In contrast, for groups of outputs, usage statistics may be obtained automatically from the hosting website API, if it has one, or from using a crawler and page scraper if not. For websites not reporting usage data, web citation counts can instead be obtained by submitting automated searches to Bing with queries designed to match web pages mentioning the outputs.

33.2.2 Indicator Robustness

Almost all online indicators for nonstandard outputs are not robust in the sense that they can be easily spammed, both deliberately and accidentally. Most of these indicators are derived from usage data, which is easy to manipulate. In the simplest case, a person may repeatedly download their own resources and a computer scientist may even write a program to repeatedly download their outputs. Some websites take steps to avoid such manipulation but it seems likely that a resourceful person could circumvent most protection measures. Accidental manipulation may take the form of uses of an output for spurious reasons. For example, a statistician may use a dataset to illustrate a statistical technique to a large class of students, asking them all to download the dataset to try out the method themselves. This would inflate the download count of the dataset and give a misleading impression of its intrinsic value.

The same is true for web citation data: someone could create many artificial web pages citing their outputs to inflate web citation counts. Thus, almost none of the data sources for nonstandard outputs are safe against manipulation. This means that these indicators should either not be used for formal evaluations or evaluators should be cautious when interpreting them. This caution could take the form of considering whether the indicators are credible in the context of other available information. For instance, an evaluator reading a researcher's claim to have created a blog with a million views might reject this evidence after visiting the blog and finding it to be poor and unlikely to have attracted much attention. In contrast, another blog might have a similar claim accepted as credible if it was clearly interesting and professional and its author had attracted a lot of media attention.

As an aside, because of the potential to manipulate most web data, Altmetric.com at the time of writing only uses data that could be tracked to its originator in its main altmetrics. For example, it includes tweet counts and reports each individual tweet but does not report Mendeley reader counts within its main outputs because they cannot be tracked to the individual readers through the Mendeley API.

The main current exceptions in terms of robust indicators for nonstandard outputs are citations to books from book citation indexes (see Chap. 27), and citations in the Clarivate Analytics Data Citation Index (DCI). The latter is an exception because the data citations originate from a (mainly) peer reviewed source: academic journal articles indexed in the Web of Science. This makes them much more difficult to manipulate and therefore reasonably credible. These exceptions use (mainly) traditional citation indexes rather than online indicators.

Given the problems with a lack of robustness for most indicators for nonstandard outputs, should they be used at all? It seems reasonable to exploit them in situations where manipulation is unlikely or pointless, such as for self-evaluations or formative evaluations [33.34], and when they can be supported by other sources of information so that an evaluator can make a judgement about their credibility. In this context, it seems desirable that for formal evaluations, indicators that could be spammed should be accompanied by an honesty declaration to state that they have not been deliberately manipulated. This would both raise awareness amongst evaluators of the possibility for manipulation and raise the stakes for those that would be prepared to dishonestly manipulate their data.

33.2.3 Indicator Evaluation

Systematic scientometric assessments of indicators for nonstandard outputs are needed to give evaluators confidence in them, even in the absence of manipulation. For citations to journal articles, a standard approach is to compare the rank order or quality categories assigned to a set of documents by citation counts with the same information produced by a set of subject experts [33.35, 36]. This uses peer review as the gold standard against which citation-based indictors should be judged. Such evaluations are rare because of the time and expense needed for subject expert judgements. There are a few exceptions, such as public evaluations of biomedical research in the F1000 website [33.37].

Citation counts are the main source of evidence used for alternative indicators for journal articles. Most evaluations have assessed the strength of the correlation with citation counts because citation counts have already been validated against peer judgements in many fields. Correlations with citation counts are therefore an indirect method of assessing value. Paradoxically, most alternative indicators are valued for their ability to highlight a different type of impact to that of citations but a positive correlation at least demonstrates that the alternative indicator is nonrandom and related in some way to scholarly activity [33.38]. Alternative methods, such as content analyses of random samples, can be used to identify the *type* of impact reflected by the indicator. Content analysis involves human judges assessing a set of texts to categorize them into coherent and relevant groups [33.39]. Content analysis and other human checking can reveal whether a potential impact indicator is reflecting a desired type of impact rather than spam [33.40, 41].

The situation is different for nonstandard output indicators because these typically do not have citation counts in Scopus or the Web of Science and so correlations with traditional citation counts cannot be calculated. Moreover, even nonstandard outputs of the same type (e. g., videos) have many different audiences and purposes and so are not homogeneous enough in impact type for a correlation test to be meaningful, even when citation data is available. Thus, systematic evaluations of nonstandard output indicators are difficult to achieve. In practice, as in the REF case study examples above, indicators are probably evaluated separately and informally for each individual output rather than collectively as a theoretical exercise.

33.3 Research Outputs for Applications

Scholars produce artifacts, such as data and software, that are designed to be exploited by others for future research or applications. Important designs or ideas may also be patented to protect their commercial value if they are used for future research. There may be associated journal articles describing these outputs but, if not, then developing impact indicators would help to give recognition to the creators.

33.3.1 Data

Projects may produce data as a natural part of their research and then publish articles that evaluate that data. Sharing the data produced would have several advantages, including some that are field-specific [33.42]. First, the analysis of the work could be checked for accuracy (reproducibility is a desired goal for science: [33.43]. Second, the data could be reanalyzed to check if the results are dependent on the analysis methods chosen (method triangulation). Third, the data could be used for other purposes, such as meta-analyses of multiple papers, or aggregated with other data for a different type of investigation.

In the past, few researchers shared their data [33.44] but there are now increasingly many incentives and funder/journal mandates to promote data sharing [33.45]. There are many free or cheap digital repositories for this purpose and an organization, DataCite, that assigns DOIs or other identifiers to data to make it easier to cite [33.46, 47]. Data sharing can have drawbacks for researchers, such as the time needed to format a dataset in a way that is suitable for sharing and the risk that other researchers will publish a study on the data that its creators had intended to do [33.48, 49]. Most researchers would be willing to use others' data, in principle, however [33.50]. For these reasons, it is important to reward researchers for data sharing by allowing them to have suitable acknowledgements of their work, such as in the form of citations or download counts. Whilst there is some evidence that sharing data associated with an article may help to attract citations to the paper [33.51], this may not be a direct enough incentive.

Although not an online indictor in the sense of the current chapter, the Clarivate Analytics DCI systematically indexes data from a large set of academic repositories and counts citations to that data from the scholarly documents that it indexes [33.52]. Its coverage is international and multidisciplinary [33.53]. It therefore gives a large and systematic source of citation counts for data that has been indexed in one of the repositories that it covers. It has broad coverage but seems to be dominated by life sciences research [33.53, 54], presumably because of extensive and systematic data sharing in this area.

Despite the large coverage of the Web of Science, about 85–89% of DCI datasets have received no citations but more recent datasets are a little more likely to be cited than older ones [33.55–57]. This may have occurred because it has become more usual to cite data used, or data re-use has become more common. Whatever the reason, most datasets will probably remain uncited in the long term but there is nevertheless still an opportunity to recognize the minority of particularly useful datasets through their citation counts. In some specific fields, however, including crystallography and genomics, data citation has become common [33.57] and appears to be central to progress [33.58].

Although citation counts are more robust, a more natural method with which to investigate the impact of data is through downloads. Each download could represent a use of the data even though few uses eventually result in a DCI citation [33.59–61]. For example, the data could be re-used for education and training [33.62]. Low positive correlations between data downloads and citations to the article originating the data suggest that a paper's data has a value that is to a large extent independent of the paper that produced it, at least for the Dryad repository [33.60]. Repositories often report download counts for their datasets that can be used as impact evidence for more typical datasets that are not in the 15% that attract DCI citations. For example, the average (geometric mean) number of downloads of datasets in FigShare is at least 11 for all subject areas as investigated in one study [33.61]. Evidence of sharing, viewing, or interacting with data may also be useful [33.63].

33.3.2 Software

Computer programs are sets of instructions for a computer to complete tasks. Software can be written in many different languages, including Java, Python, C, and Visual Basic. It can also consist of scripts to run in a software environment to draw upon and extend its existing functionality, such as Excel, Matlab, or R. Computer programs are used in research for a variety of reasons, including web data collection (e. g., web crawlers), data collection from elsewhere (e. g., movement trackers in smartphones), and data analysis. Programs can be general purpose and multifunctional (e. g., the visualization software Gephi), targeted at a field-specific task (e. g., VOSViewer for bibliometric network construction and visualization), or may complete a single task, such as a small R script to fit a hooked power law to a set of citation counts.

Software sharing can be valuable because programs can be time consuming to create and the researcher that shares software can save the time of others who would have to recreate it otherwise. Software sharing can also aid scientific quality control by allowing others to check the work presented in a publication. There are three types of software sharing. The most common type is probably to post the finished program in a repository or on its own website for others to download. Programmers may also post the source code of the program online (open source software) so that others can modify it, if necessary. Finally, a program can be shared interactively by posting it to a collaborative working environment, such as Google Code, GitHub or SourceForge, so that other programmers can collaborate on developing the code [33.64]. High-profile examples of collaboratively developed open source software include Linux, Open Office, and the web browser Mozilla [33.65]. There are many less well-known programs that are the primary work of individual academics or research groups and are in code-sharing sites. A high-profile hybrid example is the statistical software R that has a common core but can be extended by installing *packages* with additional functionality, many created by researchers [33.66].

Although there are many academic studies of the mechanics of software sharing, few have assessed whether it is reasonable to use indicators to evaluate the usefulness of the product itself. One exception found that download counts for Scopus-cited software in Google Code correlated weakly with Scopus citation counts. The low correlation was due to some code with apparently little academic value being widely used outside of academia. Unsurprisingly, software was typically downloaded many times more often than it was cited and one program had over a million downloads but only one Scopus citation [33.67]. Thus, the potential audience for a program as well as its added value to the user must be considered when evaluating software download counts. A complex and unique program may have enormous value to a few researchers that would otherwise not be able to conduct a type of research, whereas another program might be widely downloaded by the public despite the existence of similar programs from other sources, or for a trivial usage.

In addition to special-purpose software websites and ad hoc places online, code can also be stored in some general-purpose academic and other repositories. The general repository FigShare includes software as a resource type and code deposited in the site seems to have been viewed an average of at least 40 times in all research fields, with the cross-field average being 45 [33.61].

33.3.3 Patents and Products

Some scientists produce ideas or products of patentable commercial value [33.68]. Patents are documents that confer the legal right to protect an invention. They are recognized as valid research outputs in some national research assessment systems, such as that of the UK [33.69], Denmark and Italy, and the publication of a patent can be taken as an indicator of potentially useful research [33.70]. Universities may publish patents to safeguard the discoveries of their academics and so counts of patents are sometimes used as indicators of innovation or commercially relevant innovation. At the level of individual academics, the value of their patents can also be assessed financially or with the aid of citations from other patents, from scientific literature, or generally on the web. Patents can also be used to map areas of technology [33.71] and commercial technology transfer [33.72], underlining their value as indicators of commercial innovation. Patents are covered in other chapters of this book.

33.4 Multimedia Outputs

Scholars may communicate through rich visual or auditory media, including performances. The purpose may be to educate (in the pedagogical sense), inform (targeting the public or a professional audience), entertain, or culturally enrich. Some multimedia outputs are also a form of research data (e. g., diagnostic medical images, video records of animal behaviors). This section includes artistic performances even though they are not usually multimedia.

33.4.1 Presentations

Scientists routinely give presentations as part of their research. These include seminars to their own university or other institutions, talks at professional or trade conferences, and public lectures. A growing form of presentation is the video made to be posted online for later watching [33.73]. Some journals now encourage authors to record explanations of their work to post alongside the published paper in the journal website.

Scholarly presentations can be influential, but their impact is rarely tracked. Conference presentations in computing, computational linguistics, and some engineering-related fields are a partial exception [33.74, 75] because the impact of the conference proceedings paper associated with a talk can be tracked and used as primary evidence of the value of the content of the associated talk. Similarly, presentations that are associated with journal articles can also have their value subsumed into that of the article. Nevertheless, some types of research are unlikely to be eventually published in a journal [33.76] and therefore the conference presentation slides may be the main record of the information contained in them, and the only chance for the authors to track the impact of their work. This is the case for some business research areas, where PowerPoint presentations are an important outlet for scholarly knowledge [33.77].

It is increasingly common for live talks to be recorded and posted online as a video or slides with audio. The impact of scholarly videos is discussed in a different section. Presenters can also share their slides with the audience and for wider future use by posting them online. This could be on their personal website, on the conference website, or in a specialist presentation-sharing site like SlideShare [33.78]. This has created the potential to record, report and assess the impact of online presentations, as well as to use the citations from them as a new source of impact evidence for other documents [33.79].

Only one study so far seems to have explicitly assessed methods to evaluate the impact of online presentations other than videos. SlideShare is perhaps the most prominent free repository of online presentations. Although any type of presentation can be uploaded, SlideShare is owned by LinkedIn and is branded for use by the professional community, rather than scholars or educators. About 70% of recent SlideShare content consists of presentations (rather than just sets of images, for example). About 10% are of academic origin (2% associated with conferences), with the humanities, social science, business, and computing topics dominating in comparison to natural/life science, engineering, and medicine. Only a tiny minority of SlideShare presentations are cited by Scopus (by 2016, there were a total of 4436 such citations) and so there is little value in using formal citation analyses on them. Nevertheless, individual researchers could use the number of downloads per document, which averages about ten, as evidence of their presentation's impact [33.80].

33.4.2 Videos

Many academic videos are recordings of talks or lectures given to students and conferences that serve as a reminder or as an alternative means of accessing the content for people that were unable to attend. At the other extreme, some are professionally scripted, created, and edited films for online dissemination. The intended audiences for such videos might be a narrow set of scholars (e. g., for a video demonstrating specialist software or a method), a specific nonacademic group (e. g., entertaining science outreach videos aimed at school pupils), or the public.

Considering only YouTube videos that have been cited at least once in Scopus, the uses of videos vary greatly by field, but they seem to be most valuable in the arts and social sciences. The ability of a video to show a performance or film is an obvious advantage in comparison to the written word or still images. Scientific videos can also usefully demonstrate experimental procedures or surgery that would be more difficult to express in words [33.81].

As for most nonstandard academic outputs, interpreting the value of a video or set of videos is difficult because the different potential audience sizes make benchmarking or field normalization difficult. This issue is exacerbated for videos posted to general sites like YouTube, where the competitors include commercially produced popular content, such as music videos and television shows [33.82]. The task is easiest for outputs that are designed to attract a large audience because their creators would not need to explain in as much detail why the audience numbers demonstrated substantial impact—the large numbers would speak for themselves. For initiatives that produce large sets of videos, hybrid indicators have been proposed that combine the quantity of videos produced with their downloads [33.83] or Like counts and other data [33.84]. This is probably not useful in an academic context because more fine-grained information is needed to help interpret the meaning of the data.

The TED (Technology, Entertainment, Design) initiative and TED talks website are a partial exception to the difficulty of benchmarking scholarly videos:

> TED began in 1984 as a conference where Technology, Entertainment and Design converged, and today covers almost all topics—from science to business to global issues—in more than 110 languages (http://www.ted.com).

TED talks are disseminated in YouTube and the TED website, with the total number of views exceeding a billion [33.85]. A large proportion of its talks are by academics but all target a general audience and have a similar style and production input. It is therefore reasonable to benchmark the viewer figures for them against other videos from the same site to get useful impact evidence. This benchmarking should be against other videos from the same field (and of the same age) because there are broad disciplinary differences in popularity [33.85]. Of course, only a small minority of academics ever give a TED talk, and just being invited to give one is an indicator of prestige and potential public interest [33.86], although those invited seem to be a gender-, country-, and institution-biased subset of academics [33.87]. TED talks attract interactions in the form of comments from viewers, often in the form of substantial discussions, and the content of these may form a source of additional qualitative evidence about the impact of a presenter's work [33.88].

33.4.3 Images

Pictures are central to some visual arts-based scholarship and can also be useful to convey scientific information visually through graphs. In some areas of science, photographs can also be important to convey naturally visual information, such as the appearance of species, solar objects, and disease symptoms. In some of these cases the image is the scholarly output, not only in the obvious case of visual art, but also in health contexts. For instance, the *New England Journal of Medicine* accepts submissions that are images with extended captions [33.89]. Images can also be useful for science communication because visual information can be more easily communicated and is perhaps intrinsically more interesting [33.90]. Thus, for example, NASA's (National Aeronautics and Space Administration's) astronomy picture of the day archive (http://apod.nasa.gov/apod/) is a popular collection of images that seems to be widely used in schools (Fig. 33.1).

The impact of individual images within scientific articles and books does not need to be assessed separately, but there is still a need to evaluate the impact of individual images and image collections that are separate academic outputs. If these images are hosted in a standard media-sharing site, such as Flickr, then their impact can usually be assessed with the download or

Fig. 33.1 A free NASA image of Sam the monkey, recovered after a 1959 flight in an experimental spacecraft. Engaging images aid public science communication

viewing statistics provided by that site [33.91]. Like most other indicators discussed in this chapter, these would be hard to benchmark or field normalize and so it would be difficult to use them as fine-grained impact indicators.

One impact indicator is specific to images and not applicable to any other type of academic output: copying. Images that are freely shared can have their impact assessed by counting how many copies of them have been found online. Thus, a successful image would be one that could be found in many different websites and perhaps also employed for multiple different purposes, from illustrating an academic or educational article to being part of a book cover [33.92–94]. It is possible to find such image copies with Google image search and TinEye searches for image copies (i. e., search by image). Current image searching is not restricted to identical copies of an image but can also find similar images and versions that have been modified by having text added, being cropped or resized [33.93]. It does not seem to be possible to automate image searching, however, making it difficult to achieve for large collections of images and hard to gather data for benchmarking purposes. Analyzing the web pages hosting the image copies can also give useful information about the types of uses found for each image [33.94].

33.4.4 Artistic Outputs and Performances

Many artists do not write journal articles or monographs but instead produce other creative outputs, such as plays, acting, sculptures, choreography, or musical performances. For instance, *performance as research* is accepted within the arts [33.95], although it seems to be far from universal [33.96]. In the UK REF 2014, for example, artifacts, exhibitions, performances, compositions, designs, and visual media were treated as equivalent to journal articles [33.69]. A sample valid output from a high-scoring University of Manchester submission is the performance portfolio, *Reflecting on environmental change through site-based performance*, involving *Site-specific performances, in open air locations* in *Bradford* and *Bristol* [33.97]. One exhibition, part of the successful REF submission of the Courtauld Institute of Art, was *Devotion by Design: Italian Altarpieces before 1500* at the National Gallery. The context provided with this submission included:

> Devotion by design investigated ways in which altarpieces can be re-displayed to emphasise their original contexts despite being framed, literally and figuratively, by modern institutions. Additionally the exhibition presented research findings on the materiality and construction of these artefacts. [33.98]

The heterogeneity of artistic outputs makes the production of systematic data for evaluations difficult and even undesirable [33.99]. This issue is tackled by guidelines for evaluating funded arts and humanities research projects in the UK that emphasize a multiple methods approach [33.100]. Any kind of standardization of indicators can be undesirable because of the varied means through which art can influence people, sometimes in very personal and long-term ways [33.101]. Nevertheless, it seems reasonable to use an ad hoc range of methods to help evaluate artistic outputs. For example, a performance of a play might be assessed with the aid of audience numbers, contextual information about the nature of the audience (e. g., regular theater-goers versus new attendees), and the depth of the engagement with, and reaction to, the event. This information may be obtainable from questionnaires and a few in-depth interviews as well as (for larger performances) published press and academic reviews. Thus, the goal of arts evaluations may be to collect ad hoc data to help assess whether the work has had the intended effect [33.99].

Some artistic outputs can also be assessed with the aid of prestige-based indicators, such as the importance of the hosting gallery or theater or the awarding of national or international prizes relevant to the genre.

33.5 Websites

Websites may be created by academics to communicate with other scholars or the public (e. g., blogs). They may also be substantial engineering tasks to create a useful resource (e. g., digital archives).

33.5.1 Academic Websites

Researchers sometimes produce websites as an important means of communicating their ideas to a specific audience. In the UK REF 2014, websites were accepted as valid outputs, equivalent to journal articles [33.69], although both were subject to peer review for quality assessment. The work of individual scholars, research groups, departments, and institutions can also be promoted by their official websites and so a simple all-encompassing way to attempt to assess the overall impact would be to assess the popularity of the website. This basic idea led to the creation of the web impact factor (WIF), which is the number of hyperlinks pointing to a website divided by the number of pages in the website [33.102, 103]. This mimicked the Journal Impact Factor. Since hyperlinks to university websites are created for primarily academic-related reasons (including education), this seems to be a reasonable general indicator of scholarly impact [33.104–106].

It was easy to calculate this indicator for any website in the early years of the web using the hyperlink search commands of the search engines AltaVista, Microsoft Live Search and Yahoo!, but these have all now disappeared. The WIF was problematic because it penalized websites for creating many pages. If the denominator was replaced by a measure of the size of an institution then this produced a better indicator that correlated overall with the quality and amount of research produced by the institution [33.107]. Nevertheless, additional information is needed to help investigate the meaning of the link counts [33.108] so that the type of impact (if any) that they reflect can be identified.

There are several methods to assess the impact of a website other than the number of hyperlinks pointing to it. The most obvious is the number of visitors but this information is not generally shared by website owners. A practical alternative for large sites is the traffic volume ranking of Alexa.com [33.109], which gives an idea of a website's visitor traffic. It is only useful for large websites and cannot be used for subdomains of websites, such as for most individual departments and research groups. It is also possible to gather hyperlinks from part of the web, such as all universities in a single country, with a web crawler [33.107] but this is impractical for the whole web or very large websites. A practical alternative is the URL citation [33.110, 111]. This is a mention of the URL of one website in another. Thus, instead of counting the number of hyperlinks to a website, it is possible to count the number of times the website URL has been mentioned. This can be achieved by a simple search engine query. For example, to find pages mentioning any URL from the University of Wolverhampton (http://www.wlv.ac.uk/) the following Bing or Google query would work:

```
"wlv.ac.uk"
```

The query can be modified by adding an extra term to exclude pages from the University website, since these self-citations depend mainly on the size of the site.

```
"wlv.ac.uk" -site:wlv.ac.uk
```

This approach has the advantage that the data can be automatically gathered using search engine queries in Bing, such as via the Webometric Analyst free software, but the disadvantage is that there are far fewer URL citations than hyperlinks. For large websites, Bing queries return only a small fraction of the URL citations. This fraction is produced by a complex algorithm [33.24] and so the results need to be treated very cautiously. Partial solutions to the latter problem are possible by varying the search parameters to gain additional results previously hidden by Bing, either by varying the search market of a query [33.112] or by adding refining terms to generate derivative queries [33.113].

The problem that websites are typically cited by hyperlink or by name rather than by URL can be circumvented by querying Bing for mentions of the name of an organization rather than its website URL. This does not work well because many organizations have similar or derivative names and so name searches are not unique. The linked title mention method circumvents this problem through a two-stage approach, first searching for mentions of a website and then checking that the matching pages also contain a hyperlink to the relevant website [33.114].

In summary, it is possible to assess the overall impact of organizational websites as a method to assess their overall influence, combining that of their academics' standard and nonstandard outputs as well as their educational and other influences.

It may also be useful to evaluate academic-related websites on a smaller scale, such as at the level of departments, for formative evaluations or assessments of research areas. Investigations of research areas through their groups can help to identify patterns of development of new fields, as well as the relationship between key academic and nonacademic actors, at least online [33.115]. Moreover, whilst departmental or research group websites may not be often worth evaluating systematically as part of an assessment, an exception could be made when the research group's primary outputs are all nonstandard, such as in performance-based arts research areas and for institutes that generate resources to serve local industry, such as for some maritime research groups.

The websites or home pages of individual scholars could also be evaluated using URL citation counts or linked web mentions, but this again is presumably only useful in the case where an academic's reputation is their primary value and they do not have specific tangible outputs that can have their impacts assessed. Online impact assessment may be used in a formative mode to seek systematic gender, nationality, and other biases in online recognition, however [33.116]. This type of evaluation is difficult to conduct fairly because there are substantial differences in the extent to which academics engage with the web and social web [33.117, 118].

33.5.2 Digital Repositories

Digital repositories are important in some areas of scholarship. Whilst some commercial repositories have long been essential, such as newspaper archives, there have been many initiatives to digitize images, documents, maps, and other artifacts both as a means of preserving them and to reach a wide audience. Examples include the Lives of the First World War archive of photographs, letters, and other digital artifacts (livesofthefirstworldwar.org), the Internet Archive's moving image archive (archive.org/details/movies), and the British Library's historical British Newspaper Archive (http://www.britishnewspaperarchive.co.uk). Some universities, such as the UK's University for the Creative Arts (http://www.research.ucreative.ac.uk), also have specialist digital archives to record or document artistic outputs, including performances [33.119]. Digital repositories can be substantial outputs produced by teams of academics, librarians, and computer scientists and may even be their primary outputs over a period of years during the development phase.

Whilst a logical way to assess the value of a digital repository is by counting visitors or downloads, they are often too important to employ such a basic indicator [33.120]. Moreover, usage information is insufficient due to the lack of benchmarking data because usage data is rarely shared by repositories. Because of these problems a suite of methods has been recommended for evaluations. This should include usage data but also web citation data (e.g., URL citations) because these can be benchmarked against comparable repositories, if any. Interviews with end users, log file analysis, and content analyses of URL citations are recommended to get more detailed insights into the nature and depth of the value generated by each repository [33.121, 122]. Web citation data on comparable repositories can point to types of uses that are possible for the repository being evaluated, even if they are not occurring for it—in other words, missed opportunities [33.123]. Additional resource-specific methods may also be possible in some cases, such as reverse image lookup to find copies of images shared from image archives [33.94].

Fine-grained web server log file analyses may be helpful to discover more about a repository's visitors. Each user of a website must supply information about their location to the web server to communicate with it and this is routinely logged for monitoring purposes. Analyzing such web server log files can then reveal the origins of the visitors in broad terms, such as whether they are likely to be accessing from academic, governmental, personal, or commercial premises. Aggregating this information can point to the sectors of society in which a repository has the most impact, in addition to geographic information, such as country of origin [33.124]. This information is imperfect because a person's location does not necessarily reflect their activity type—such as people working from home—but can give deeper insights than simple usage data. This is helpful even for infrastructures that get extensively cited in formal academic publications because it can point to otherwise overlooked user groups [33.125].

Despite the proposals discussed above, it is easy to overlook the benefits of a repository for unknown user groups. Thus, there is an ongoing need to develop methods to give insights into the value of digital repositories—and other expensive research infrastructures [33.120].

33.5.3 Blogs

Blogs are simple websites that center on a collection of separate posts that are displayed in reverse chronological order. Many scholars maintain a blog to discuss aspects of their research. Whilst in many cases blogs are secondary outputs, some are widely read and reliably disseminate information to a large audience. Important blogs include The Impact Blog of the London School of Economics [33.126], which contains content mainly about social sciences research impact that has been authored by many different scholars, and *DC's Improbable Science*, a blog created by a single biochemist that claims over three million views and includes many discussions on science, including many attacks on bibliometrics and altmetrics [33.127]. If a blog is an important output of a scholar or group of scholars, then it would be useful to be able to evaluate its impact with the help of appropriate indicators.

Research blogs typically discuss published research and often focus on new journal articles that seem to be important, such as due to publication in a prestigious journal [33.128]. Their role is not just to discuss and evaluate research but also to translate it for a lay audience or nonspecialist researchers, particularly in the life sciences [33.129–132]. Blogs may reach a wider audience than the articles reviewed [33.133] because of these goals.

Although there is much research about blogs [33.134], there do not seem to have been systematic studies of blog impact, other than individual case studies [33.135]. The logical choice for a blog impact indicator is the number of visitors or the number of page views. One UK REF case study reported [33.136]:

> An outreach campaign has communicated SCCS's (School of Contemporary Chinese Studies) economic research to a broader audience. Yao's blog [h], first launched in April 2010 by Beijing-based finance and economics publication Caijing, has now attracted more than a million hits.

And from another [33.137]:

> Butterworth's *Life and Physics* blog [] attracts a sustained average of about 50 000 unique visitors a month, with peaks of around 20 000 a day for key posts at key times.

Thus, despite the absence of benchmarks and supporting scientometric research, simple blog indicators are being used to help demonstrate the impact of blogs as nonstandard research outputs.

33.6 Documentary Outputs

Whilst journal articles, conference papers, and books are discussed elsewhere in this volume and may naturally be evaluated with traditional citation indexes, grey literature and dissertations are important in some fields. Grey literature may target nonacademic audiences and so it is natural to seek online evidence of its impact. Dissertations may be important in the humanities when a thesis is not subsequently published as a monograph.

33.6.1 Grey Literature

The term grey literature refers to documents that are informally circulated online or offline rather than as an officially published book or as part of an edited volume, academic journal, or other recognized publication venue. Many grey literature documents are freely shared online in the form of PDF or Microsoft Word documents although they may also exist as web pages.

Grey literature is extensively cited in health research, including technical reports prepared by international organizations like the United Nations [33.138]. Traditional citation analyses can be conducted on grey literature, although the citation counts are not likely to be high. Google Scholar is an appropriate tool for this because it also indexes some grey literature [33.139], but perhaps not all [33.140]. Microsoft Academic is an alternative that allows automatic data collection [33.141]. Nevertheless, most technical reports of this type are not aimed at scholars and so academic citation counts would give a misleading impression of their level of uptake. In veterinary medicine, citations to the grey literature are common, accounting for about 6% of all journal article references (this figure includes conference papers as grey literature, accounting for half of the 6%), although the proportion varies substantially by specialism [33.142]. Excluding conference papers, most of the grey literature originated from governmental or commercial organizations. Research reports are also important in a policy context. A survey of senior UK civil servants found that they accessed academic expertise from these more often than from journal articles, books, or any other publication type [33.143].

Grey literature is an important output of some research organizations. A study of the outputs of a marine advisory body found that two thirds of its Web of Science citations were to its reports rather than its monographs and journal articles [33.144]. This shows that traditional citation analysis can give nontrivial results for some types of grey literature. In medicine and some other areas of research it is important to publish unsuccessful studies or those without statistically significant findings to ensure that future meta analyses and systematic reviews have findings that are unduly influenced by publication bias [33.145, 146].

Informally published documents can be posted anywhere on the web but scientific repositories offer enhanced visibility and long-term preservation. The preprint archive arXiv [33.147], for instance, contains many otherwise unpublished documents, as do RePEc (Research Papers in Economics; economics) and SSRN (Social Science Research Network; social sciences). Although these are discipline-based they still attract citations from outside of their home discipline [33.148] and can therefore be used for the wider dissemination of research. Whilst some reports in archives are rejected journal articles, others are research reports that were not designed for peer review. Unlike for arXiv, grey literature posted to FigShare has associated usage statistics within the site [33.61] that can be used to track impact. RePEc (Research Papers in Economics) has an extensive suite of statistics, including the h-index and others aggregated at the level of authors (and institutions) in addition to paper-level statistics. Author-level indicators with RePEc data are imperfect [33.149], for example due to missing citations, and so RePEc indicators should be used cautiously. RePEc also includes measures to detect and exclude manipulation and so its data is probably more robust than that of most other repositories [33.150]. At the time of writing, for each uploaded paper, SSRN (Social Science Research Network) provided the number of downloads and abstract views as well as the rank within the repository based on the download count (for an alternative ranking, [33.151]). Thus, if authors wish to track the impact of their informally published reports, then it would help to deposit them in a place that tracks their usage.

An alternative strategy, and a method to obtain deeper insights into the use of grey literature, is the Web impact report (WIRe) [33.152]. This includes a range of types of online impact evidence for a collection of grey literature and benchmarks it against similar sets produced by other organizations. The basic method to identify online impact evidence is a web citation search. This is a commercial search engine query for each document by name and author to identify how often it has been cited online. The use of web searches rather than Scopus of WoS citations stems from the assumption that much grey literature targets wider audiences than publishing academics. Thus, Googling with the query below would match web pages mentioning the NatPaCT leaflet *Ten Steps to SMART objectives*.

```
"Ten Steps to SMART objectives
natpact"
```

The query below is an improvement by excluding self-citations.

```
"Ten Steps to SMART objectives
natpact" -site:natpact.info
```

This gives a simple web citation count for each document. The citing URLs can then be broken down by top-level domain (TLD) and the national TLDs (e. g., .uk, .de) would give evidence of international spread. A parallel content analysis of the citing pages is recommended to give deeper evidence about who was using the grey literature documents and how they were using it. The use of web citations from search engines is limited by the partial web coverage of search engines, as discussed above, but also because grey literature documents may be cited informally online. Informal citations may omit the author, shorten the document name, or even just describe it as a report from a given organization.

33.6.2 Dissertations

The PhD thesis is a nonstandard academic output in the sense discussed here, although produced by most academics at the start of their careers. The thesis is a book-like document (see the chapter on book impact assessment) and theses may contain material that is not otherwise published. Although science PhDs may attempt to publish their key findings in journal articles, humanities scholars may turn their dissertations into monographs, and some countries require PhDs to be published as books (e. g., by university presses), it still seems likely that this is not the general rule. Thus, especially for junior scholars, it may be useful to attempt to assess the impact of their dissertations. The increasingly common institutional requirement to publish dissertations electronically online [33.153] help evaluations by publicizing them online and simplifying access.

The quality of dissertations or the success of the scholars that produced them may be outputs for the main supervisors [33.154]. Whilst numbers of successful PhD supervisions are already seen as valid outputs in some research assessment exercises (e. g., New Zealand, UK), academics may list individual students supervised in their CVs for more fine-grained evidence. Evidence from library and information science suggests that successful doctoral student mentoring is a dimension of scholarship that is not well reflected by citation counts and is therefore valuable to assess separately [33.155].

The primary audience of most dissertations is probably other scholars from the same field, rather than any section of the wider public. Dissertations are long documents that need to satisfy examiners and are likely to be complex and difficult for nonspecialists to read. Thus, traditional citation counts are the logical choice of indicator for theses. Nevertheless, the similarity between theses and books and the scholarly connection in some fields [33.156, 157] suggest that citations from books (via Google Books, Scopus, or WoS) should be obtained in addition to citations from journal articles, if possible. The ProQuest dissertation database gives an additional source of impact evidence for the theses indexed in it. It originates from the USA but has international coverage and tracks the usage of the dissertation in its collection [33.158].

33.7 Reputation

An important aspect of a scholar's profile is their reputation inside and outside of academia. This can be thought of as a nonstandard output of their scholarly work. The h-index is an obvious (but flawed) reputation indicator, and is covered in another chapter.

Reputation can be assessed informally or with the aid of a CV that lists ad hoc indicators of prestige. These may include invitations to give talks and keynote presentations, as well as editorial positions and awards [33.159, 160]. Invitations can apply to all levels of achievement. Even a PhD student might get asked to talk to a departmental seminar, for instance. Depending on the field of scholarship, an academic might be asked to exhibit their work in a prestigious venue, accept a valued national training role, speak on radio or television, or give evidence to a parliamentary select committee. Scholars may also apply for prestigious roles, such as association president or committee chair.

Reputation could also be assessed by counting links to an academic's home page or by measuring their followers in sites that they use. This would be unfair for sites like Twitter, that often combine social and informational elements [33.161]. Whilst Academia.edu and ResearchGate are both natural places to evaluate reputation [33.162], not all academics use these sites with some disciplines being more active users than others and newer articles being more likely to be registered [33.163]. The usage and reputation scores in these academic sites probably reflect use of the site to a large extent, rather than overall reputation [33.164]. There is evidence of ResearchGate scores for individuals primarily reflecting activity within the site rather than external evidence of success, for example [33.165–167]. Like Twitter, these sites can merge personal and professional interactions [33.165, 168]. Nevertheless, reputation in Academica.edu seems to be primarily driven by academic rather than personal factors [33.169], although more evidence is needed to confirm this and it might change over time as the users and uses of the site evolve. At the institutional level, ResearchGate scores seem to reflect university rankings from other sources, although there are big national differences in uptake for the site [33.170].

Influence is like reputation because an influential academic would presumably have a good reputation and a highly regarded scholar would have the potential to be influential. There has been much research to develop methods to assess influence in social media [33.171], for example as reflected in the likelihood that someone's tweets get extensively retweeted. Although this type of influence does not seem to be widely recognized as a useful academic contribution, it may be a useful indicator of reputation. Thus, academics in the future might like to quantify their (relevant) social media influence to claim an online reputation. A similar concept is *authority* in the sense of being a recognized source of information on a given topic. This too can be quantified online [33.172].

One study analyzed speaking fees charged by some scholars, finding (indirect) evidence that this is a reasonable indicator of their reputation outside of academia and that this is probably not directly influenced by their reputation inside academia [33.86]. Thus, speaking fees could be an indicator of the societal reputation of scholars, although only for the minority that charge them, and this information is not available online.

A more general way to assess the reputation of individual academics is to count how often they have been mentioned on the web, perhaps filtering out self-mentions from their host institutions and publications [33.173]. Depending on the scholar, such online mentions may include media coverage, conference presentations, and invited talks. The main drawback is that human names are rarely unique and so extensive manual filtering would be needed for most individuals and searching the web for mentions of some scholars, such as any of the computer scientists called Michael Jackson, would be very difficult. Nevertheless, in some arts and humanities fields, reputation is the main (intangible) output of some scholars and so counts of web mentions could help in their evaluations.

33.8 Summary: The Importance of Context

As should be evident from the different sections of this chapter, the many different types of output produced by scholars can make contributions either directly to scholarship or to the wider goals of science, including infrastructure and outreach or dissemination activities. Some are the main products of the scholars concerned and without effective ways to evaluate their contribution to academia the researchers are likely to be undervalued and marginalized within science. Currently, most of these activities are probably evaluated only through peer judgement in appointment, promotion, and funding decisions and it would be helpful to be able to provide some form of supporting quantitative data.

Quantitative indicators to support evaluations of nonstandard academic outputs rarely include traditional citation counts because most of the outputs discussed here are not likely to be cited in the academic literature. The main exception is the dissertation, and, to some extent the grey literature, but there is also an increasing move to encourage the formal citation of datasets and software. For most other types of output, usage data, when available, or web citation counts (i. e., counts of how often they have been mentioned online) are reasonable substitutes. They can also be additional sources of evidence for artifacts that are traditionally cited by revealing nonacademic uses that generate web citations.

Except for dissertations, interpreting the numbers generated by web citation or usage data is difficult. The variety of types of output, audience size, and intended purpose make it hard to benchmark scores to judge them high or low unless the numbers are extreme enough for this to be self-evident. An important issue here (discussed explicitly for UK REF case studies) is that both depth and breadth of engagement are desirable [33.174], so the exact number of uses may not reveal the amount of impact. A resource could be used by many people for trivial purposes or extensively by a small group of specialists that rely upon it for their work. Thus, in most cases the solution is to provide context to the numbers. The most natural way to do this—and again often evident in REF impact case studies and recommended in the ACUMEN portfolio—is to include a narrative that makes an explicit claim for a type of impact and then uses multiple sources of quantitative and qualitative evidence to back up that claim.

A generic problem for most online indicators is that they are easily manipulated, either accidentally or deliberately, and are therefore difficult to rely upon in formal evaluations. Thus, if they are reported for formal evaluations (e. g., end of project reports), then evaluators should not take them at face value but should critically analyze their reliability and look for multiple sources of evidence (as mentioned above) and their own common sense to cross-check and assess the credibility of any claim. Honesty declarations may also help.

For high-value resources, such as digital repositories or grey literature collections that are major outputs from an organization, it may be feasible to go further than gathering numbers by using complementary approaches, such as a human content analysis of citing sources or user interviews and questionnaires to get deeper insights into the users and uses. This may take the form of using a formal digital toolkit or an ad hoc collection of discipline-specific [33.23] or appropriate generic methods.

As a final warning, although the use of indicators for nonstandard outputs is a positive step towards recognizing diverse contributions to academia, there are informal contributions that are too difficult to quantify in practice, such as informal mentoring [33.175] and any systematic use of indicators should be aware of the likelihood of introducing biases, including probably gender biases, against those who conduct activities with impacts that are difficult to record.

Despite the studies discussed here, evaluating nonstandard outputs is an under-researched area within scientometrics and in the increasing culture of evaluating academics for their scholarly activities the lack of adequate research risks a situation in which researchers may avoid types of activity for which they will not be recognized. This has been alleviated to some extent in the UK by the mandatory REF impact case studies [33.176], which almost require research groups to engage outside academia and to consider nonstandard outputs, but it is nevertheless a potential threat to the smooth working of science on a global scale.

References

33.1 T.C. Engels, T.L. Ossenblok, E.H. Spruyt: Changing publication patterns in the social sciences and humanities, Scientometrics **93**(2), 373–390 (2012)

33.2 D. Torres-Salinas, N. Robinson-Garcia, Á. Cabezas-Clavijo, E. Jiménez-Contreras: Analyzing the citation characteristics of books: Edited books, book series and publisher types in the book citation index, Scientometrics **98**(3), 2113–2127 (2014)

33.3 ACUMEN: Portfolio, http://research-acumen.eu/portfolio (2015)

33.4 ACUMEN: Portfolio, http://research-acumen.eu/wp-content/uploads/Blank_AcumenPortfolio.v13x.pdf (2015)

33.5 T. Greenhalgh, N. Fahy: Research impact in the community-based health sciences: An analysis of 162 case studies from the 2014 UK research excellence framework, BMC Med. **13**(1), 232 (2015)

33.6 G.N. Samuel, G.E. Derrick: Societal impact evaluation: Exploring evaluator perceptions of the characterization of impact under the REF2014, Res. Evaluation **24**(3), 229–241 (2015)

33.7 Research Excellence Framework: Impact case studies, http://impact.ref.ac.uk/CaseStudies (2014)
33.8 Research Excellence Framework: Impact case study "Rublack", http://impact.ref.ac.uk/CaseStudies/CaseStudy.aspx?Id=23345 (2014)
33.9 Research Excellence Framework: Impact case study "Million-pound donors: Shaping policy and professional practice in philanthropy and fundraising from high net worth individuals", http://impact.ref.ac.uk/CaseStudies/CaseStudy.aspx?Id=4607 (2014)
33.10 Research Excellence Framework: Impact case study "10 minute puzzle podcasts", http://impact.ref.ac.uk/CaseStudies/CaseStudy.aspx?Id=43372 (2014)
33.11 Research Excellence Framework: Impact case study "Impact of machine-learning based visual analytics", http://impact.ref.ac.uk/CaseStudies/CaseStudy.aspx?Id=37006 (2014)
33.12 Research Excellence Framework: Impact case study "Mrs. Peabody investigates: Enhancing public understanding of German, European and international crime fiction", http://impact.ref.ac.uk/CaseStudies/CaseStudy.aspx?Id=5904 (2014)
33.13 Research Excellence Framework: Impact case study "Communicating research to the public through YouTube", http://impact.ref.ac.uk/CaseStudies/CaseStudy.aspx?Id=32739 (2014)
33.14 Research Excellence Framework: Impact case study "Preventing disease through promotion of handwashing with soap", http://impact.ref.ac.uk/CaseStudies/CaseStudy.aspx?Id=41463 (2014)
33.15 M. Thelwall: Weak benchmarking indicators for formative and semi-evaluative assessment of research, Res. Evaluation **13**(1), 63–68 (2004)
33.16 L. Waltman, N.J. van Eck, T.N. van Leeuwen, M.S. Visser, A.F. van Raan: Towards a new crown indicator: Some theoretical considerations, J. Informetr. **5**(1), 37–47 (2011)
33.17 L. Björneborn, P. Ingwersen: Toward a basic framework for webometrics, J. Am. Soc. Inf. Sci. Technol. **55**(14), 1216–1227 (2004)
33.18 K.J. Holmberg: *Altmetrics for Information Professionals: Past, Present and Future* (Chandos, Oxford 2015)
33.19 J. Priem, D. Taraborelli, P. Groth, C. Neylon: Altmetrics: A manifesto, http://altmetrics.org/manifesto/ (2010)
33.20 K. Kousha, M. Thelwall: Web indicators for research evaluation, part 3: Books and non-standard outputs, Profesion. Inf. **24**(6), 724–736 (2015)
33.21 M. Thelwall, D. Stuart: Web crawling ethics revisited: Cost, privacy and denial of service, J. Am. Soc. Inf. Sci. Technol. **57**(13), 1771–1779 (2006)
33.22 J. Bar-Ilan: Methods for measuring search engine performance over time, J. Am. Soc. Inf. Sci. Technol. **53**(4), 308–319 (2002)
33.23 T. Greenhalgh, J. Raftery, S. Hanney, M. Glover: Research impact: A narrative review, BMC Med. **14**(1), 78 (2016)
33.24 M. Thelwall, P. Sud: Webometric research with the Bing Search API 2.0, J. Inf. **6**(1), 44–52 (2012)
33.25 A. Uyar: Google stemming mechanisms, J. Inf. Sci. **35**(5), 499–514 (2009)
33.26 D. Lewandowski, H. Wahlig, G. Meyer-Bautor: The freshness of web search engine databases, J. Inf. Sci. **32**(2), 131–148 (2006)
33.27 L. Vaughan, M. Thelwall: Search engine coverage bias: Evidence and possible causes, Inf. Process. Manag. **40**(4), 693–707 (2004)
33.28 M. Thelwall: *Web Indicators for Research Evaluation: A Practical Guide* (Morgan Claypool, San Rafael 2017)
33.29 M. Thelwall: Are Mendeley reader counts high enough for research evaluations when articles are published?, Aslib J. Inf. Manag. **69**(2), 174–183 (2017)
33.30 A.W. Harzing, S. Alakangas: Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison, Scientometrics **106**(2), 787–804 (2016)
33.31 M. Thelwall: Microsoft Academic automatic document searches: Accuracy for journal articles and suitability for citation analysis, J. Inf. **12**(1), 1–9 (2018)
33.32 M. Thelwall: Does Microsoft Academic find early citations?, Scientometrics **114**(1), 325–334 (2018), https://doi.org/10.1007/s11192-017-2558-9
33.33 M. Thelwall, K. Kousha: ResearchGate versus Google Scholar: Which finds more early citations?, Scientometrics **112**(2), 1125–1131 (2017)
33.34 P. Wouters, R. Costas: Users, narcissism and control: Tracking the impact of scholarly publications in the 21st century. Utrecht: SURF foundation, http://www.surffoundation.nl/nl/publicaties/Documents/Users%20narcissism%20and%20control.pdf (2012)
33.35 G. Abramo, C.A. D'Angelo, F. Di Costa: National research assessment exercises: A comparison of peer review and bibliometrics rankings, Scientometrics **89**(3), 929 (2011)
33.36 J. Wilsdon, L. Allen, E. Belfiore, P. Campbell, S. Curry, S. Hill, J. Hill: The metric tide: Report of the independent review of the role of metrics in research assessment and management, http://www.hefce.ac.uk/pubs/rereports/Year/2015/metrictide/Title,104463,en.html (2015)
33.37 E. Mohammadi, M. Thelwall: Assessing non-standard article impact using F1000 labels, Scientometrics **97**(2), 383–395 (2013)
33.38 P. Sud, M. Thelwall: Evaluating altmetrics, Scientometrics **98**(2), 1131–1143 (2014)
33.39 K.A. Neuendorf: *The Content Analysis Guidebook* (SAGE, Oxford 2016)
33.40 M. Thelwall, A. Tsou, S. Weingart, K. Holmberg, S. Haustein: Tweeting links to academic articles, Cybermetrics **17**(1) (2013), http://cybermetrics.cindoc.csic.es/articles/v17i1p1.html
33.41 A. Tsou, T. Bowman, A. Ghazinejad, C.R. Sugimoto: Who tweets about science? In: *Proc. 15th Int. Soc. Sci. Inf. Conf. (ISSI2015)*, ed. by A.A. Salah,

Y. Tonta, A.A.A. Salah, C. Sugimoto, U. Al (Boğaziçi Univ. Printhouse, Istanbul 2015) pp. 28–36

33.42 O.J. Reichman, M.B. Jones, M.P. Schildhauer: Challenges and opportunities of open data in ecology, Science **331**(6018), 703–705 (2011)

33.43 P. Vandewalle, J. Kovacevic, M. Vetterli: Reproducible research in signal processing, IEEE Signal Process. Mag. **26**(3), 37–47 (2009)

33.44 D.S. Caetano, A. Aisenberg: Forgotten treasures: The fate of data in animal behaviour studies, Animal Behav **98**(1), 1–5 (2014)

33.45 A. Dinsmore, L. Allen, K. Dolby: Alternative perspectives on impact: The potential of ALMs and altmetrics to inform funders about research impact, PLoS Biol. **12**(11), e1002003 (2014)

33.46 J. Brase, I. Sens, M. Lautenschlager: The tenth anniversary of assigning DOI names to scientific data and a five year history of data cite, D-Lib Mag. (2015), https://doi.org/10.1045/january2015-brase

33.47 N. Robinson-Garcia, P. Mongeon, W. Jeng, R. Costas: Evaluating the possibilities of DataCite for developing 'Open Data metrics' on the production and usage of datasets worldwide. In: *3AM Conf* (2016), http://wdb.ugr.es/~elrobin/wp-content/uploads/3AM_summary_final.pdf

33.48 C.L. Borgman: The conundrum of sharing research data, J. Am. Soc. Inf. Sci. Technol. **63**(6), 1059–1078 (2012)

33.49 X. Huang, B.A. Hawkins, F. Lei, G.L. Miller, C. Favret, R. Zhang, G. Qiao: Willing or unwilling to share primary biodiversity data: Results and implications of an international survey, Conserv. Lett. **5**(5), 399–406 (2012)

33.50 C. Tenopir, S. Allard, K. Douglass, A.U. Aydinoglu, L. Wu, E. Read, M. Frame: Data sharing by scientists: Practices and perceptions, PLoS One **6**(6), e21101 (2011)

33.51 H.A. Piwowar, R.S. Day, D.B. Fridsma: Sharing detailed research data is associated with increased citation rate, PLoS One **2**(3), e308 (2007)

33.52 Clarivate Analytics: The data citation index, http://wokinfo.com/products_tools/multidisciplinary/dci/ (2015)

33.53 L.L. Pavlech: Data citation index, J. Med. Libr. Assoc. **104**(1), 88–90 (2016)

33.54 D. Torres-Salinas, A. Martín-Martín, E. Fuente-Gutiérrez: An introduction to the coverage of the Data Citation Index (Thomson-Reuters): Disciplines, document types and repositories, Rev. Esp. Doc. Cient. **37**(1), e036 (2014)

33.55 D. Torres-Salinas, E. Jiménez-Contreras, N. Robinson-García: How many citations are there in the Data Citation Index? Proceedings of the STI Conference, Leiden, The Netherlands. Online at http://arxiv.org/abs/1409.0753. Accessed on 1 January 2015

33.56 I. Peters, P. Kraker, E. Lex, C. Gumpenberger, J. Gorraiz: Research data explored: An extended analysis of citations and altmetrics, Scientometrics **107**(2), 723–744 (2016)

33.57 N. Robinson-García, E. Jiménez-Contreras, D. Torres-Salinas: Analyzing data citation practices using the data citation index, J. Assoc. Inf. Sci. Technol. **67**(12), 2964–2975 (2016)

33.58 P. Anagnostou, M. Capocasa, N. Milia, G.D. Bisol: Research data sharing: Lessons from forensic genetics, Forensic Sci. Int. Genet. **7**(6), e117–e119 (2013)

33.59 L. He, V. Nahar: Reuse of scientific data in academic publications: An investigation of Dryad Digital Repository, Aslib J. Inf. Manag. **68**(4), 478–494 (2016)

33.60 M. Thelwall, K. Kousha: Do journal data sharing mandates work? Life sciences evidence from dryad, Aslib J. Inf. Manag. **69**(1), 36–45 (2016)

33.61 M. Thelwall, K. Kousha: Figshare: A universal repository for academic resource sharing?, Online Inf. Rev. **40**(3), 333–346 (2016)

33.62 J.C. Wallis, E. Rolando, C.L. Borgman: If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology, PLoS One **8**(7), e67332 (2013)

33.63 P. Ingwersen, V. Chavan: Indicators for the data usage index (DUI): An incentive for publishing primary biodiversity data through global information infrastructure, BMC Bioinform **12**(Suppl 15), S3 (2011)

33.64 K.R. Lakhani, E.V. Hippel: How open source software works: "Free" user-to-user assistance, Res. Policy **32**(6), 923–943 (2003)

33.65 A. Mockus, R.T. Fielding, J.D. Herbsleb: Two case studies of open source software development: Apache and mozilla, ACM Trans. Softw. Eng. Methodol. **11**(3), 309–346 (2002)

33.66 C. Calenge: The package adehabitat for the R software: A tool for the analysis of space and habitat use by animals, Ecol. Model. **197**(3), 516–519 (2006)

33.67 M. Thelwall, K. Kousha: Academic software downloads from Google Code: Useful usage indicators?, Inf. Res. **21**(1), 709 (2016)

33.68 P.Y.H. Lee: Inverting the logic of scientific discovery: Applying common law patentable subject matter doctrine to constrain patents on biotechnology research tools, Harvard J. Law Technol. **19**(1), 79–109 (2005)

33.69 Research England: Research Excellence Framework, https://www.ref.ac.uk/guidance/

33.70 M. Meyer: Academic patents as an indicator of useful research? A new approach to measure academic inventiveness, Res. Evaluation **12**(1), 17–27 (2003)

33.71 S. Xu, X. Wang, Z. Liu, C. Luan: Network structural analysis of technology: A study from patent perspective, J. Sci. Technol. Policy China **4**(3), 214–235 (2013)

33.72 Z. Qu, X. Shen, K. Ding: Comparative analysis on technologies between Chinese and American large-sized oil companies based on patentometrics, J. Scientometr. Res. **3**(2), 68–74 (2014)

33.73 S.B. Johnson: On-line pesticide training with narrated PowerPoint presentations, J. Ext. **53**(3) (2015), http://www.joe.org/joe/2015june/tt8.php

33.74 C. Lisée, V. Larivière, É. Archambault: Conference proceedings as a source of scientific information: A bibliometric analysis, J. Am. Soc. Inf. Sci. Technol. **59**(11), 1779–1784 (2008)

33.75 M. Oppermann: Tourism conferences—academic titillation, social interactions or job market?, Tour. Manag. **18**(5), 255–257 (1997)

33.76 M. Petticrew, M. Egan, H. Thomson, V. Hamilton, R. Kunkler, H. Roberts: Publication bias in qualitative research: What becomes of qualitative research presented at conferences?, J. Epidemiol. Community Health **62**(6), 552–554 (2008)

33.77 D. Schoeneborn: The pervasive power of Power Point: How a genre of professional communication permeates organizational communication, Organ. Stud. **34**(12), 1777–1801 (2015)

33.78 P. Eveillard: Social networks in 10 lessons: No. 3 entrust your presentations with slide share!, Rev. Prat. **63**(3), 384 (2013)

33.79 M. Thelwall, K. Kousha: Online presentations as a source of scientific impact? An analysis of power point files citing academic journals, J. Am. Soc. Inf. Sci. Technol. **59**(5), 805–815 (2008)

33.80 M. Thelwall, K. Kousha: SlideShare presentations, citations, users and trends: A professional site with academic and educational uses, J. Assoc. Inf. Sci. Technol. **68**(8), 1989–2003 (2017)

33.81 K. Kousha, K. Thelwall, M. Abdoli: The role of online videos in research communication: A content analysis of YouTube videos cited in academic publications, J. Am. Soc. Inf. Sci. Technol. **63**(9), 1710–1727 (2012)

33.82 B. Haran, M. Poliakoff: The periodic table of videos, Science **332**(6033), 1046–1047 (2011)

33.83 R. Hovden: Bibliometrics for internet media: Applying the h-index to YouTube, J. Am. Soc. Inf. Sci. Technol. **64**(11), 2326–2331 (2013)

33.84 C. Xiao, Y. Xue, Z. Li, X. Luo, Z. Qin: Measuring user influence based on multiple metrics on YouTube. In: *7th Int. Symp. Parallel Archit. Algorithms Program. (PAAP2015)* (IEEE, Menlo Park 2015) pp. 177–182

33.85 C.R. Sugimoto, M. Thelwall: Scholars on soap boxes: Science communication and dissemination in TED videos, J. Am. Soc. Inf. Sci. Technol. **64**(4), 663–674 (2013)

33.86 H.F. Chan, B.S. Frey, J. Gallus, M. Schaffner, B. Torgler, S. Whyte: Do the best scholars attract the highest speaking fees? An exploration of internal and external influence, Scientometrics **101**(1), 793–817 (2014)

33.87 C.R. Sugimoto, M. Thelwall, V. Larivière, A. Tsou, P. Mongeon, B. Macaluso: Scientists popularizing science: Characteristics and impact of TED talk presenters, PLoS One **8**(4), e62403 (2013)

33.88 A. Tsou, M. Thelwall, P. Mongeon, C. Sugimoto: A community of curious souls: An analysis of commenting behavior on TED talks videos, PLoS One **9**(4), e93609 (2014)

33.89 The New England Journal of Medicine: Author center, images in clinical medicine, https://www.nejm.org/author-center/images-in-clinical-medicine (2018)

33.90 K. Schröder, H. Lüthen: Astrophotography. In: *Handbook of Practical Astronomy*, ed. by G.D. Roth (Springer, Berlin 2009) pp. 133–173

33.91 E. Angus, M. Thelwall, D. Stuart: Flickr's potential as an academic image resource: An exploratory study, J. Librariansh. Inf. Sci. **42**(4), 268–278 (2010)

33.92 I. Kirton, M. Terras: Digitisation and dissemination: A reverse image lookup study to assess the reuse of images of paintings from the National Gallery's website, J. Digit. Humanit. **3**(1) (2014), http://journalofdigitalhumanities.org/3-1/reverse-image-lookup-paintings-digitisation-reuse/

33.93 K. Kousha, M. Thelwall, S. Rezaie: Can the impact of scholarly images be assessed online? An exploratory study using image identification technology, J. Am. Soc. Inf. Sci. Technol. **61**(9), 1734–1744 (2010)

33.94 M. Reilly, S. Thompson: Reverse image lookup: Assessing digital library users and reuses, J. Web Librariansh. **11**(1), 56–68 (2017)

33.95 S. Riley, L. Hunter (Eds.): *Mapping Landscapes for Performance as Research: Scholarly Acts and Creative Cartographies* (Springer, Berlin 2009)

33.96 C.D. Brown: Straddling the humanities and social sciences: The research process of music scholars, Libr. Inf. Sci. Res. **24**(1), 73–94 (2002)

33.97 Research Excellence Framework: Output details "Reflecting on environmental change through site-based performance", http://results.ref.ac.uk/Submissions/Output/2863017?searchId=437searchPageNumber=1 (2014)

33.98 Research Excellence Framework: Output details 'Devotion by design: Italian Altarpieces before 1500 (National gallery company)', http://results.ref.ac.uk/Submissions/Output/74860?searchId=438searchPageNumber=1 (2014)

33.99 M. Thelwall, M. Delgado: Arts and humanities research evaluation: No metrics please, just data, J. Doc. **71**(4), 817–833 (2015)

33.100 AHRC: Understanding your project: A guide to self-evaluation, http://www.ahrc.ac.uk/documents/guides/understanding-your-project-a-guide-to-self-evaluation/ (2017)

33.101 E. Belfiore, O. Bennett: Beyond the "Toolkit Approach": Arts impact evaluation research and the realities of cultural policy-making, J. Cult. Res. **14**(2), 121–142 (2010)

33.102 P. Ingwersen: The calculation of web impact factors, J. Doc. **54**(2), 236–243 (1998)

33.103 A.G. Smith: A tale of two web spaces: Comparing sites using web impact factors, J. Doc. **55**(5), 577–592 (1999)

33.104 J. Bar-Ilan: A microscopic link analysis of academic institutions within a country—the case of Israel, Scientometrics **59**(3), 391–403 (2004)

33.105 J. Bar-Ilan: What do we know about links and linking? A framework for studying links in academic environments, Inf. Process. Manag. **41**(4), 973–986 (2005)

33.106 D. Wilkinson, G. Harries, M. Thelwall, L. Price: Motivations for academic Web site interlinking: Evidence for the Web as a novel source of information on informal scholarly communication, J. Inf. Sci. **29**(1), 49–56 (2003)
33.107 M. Thelwall, G. Harries: Do better scholars' web publications have significantly higher online impact?, J. Am. Soc. Inf. Sci. Technol. **55**(2), 149–159 (2004)
33.108 M. Thelwall: Interpreting social science link analysis research: A theoretical framework, J. Am. Soc. Inf. Sci. Technol. **57**(1), 60–68 (2006)
33.109 L. Vaughan, R. Yang: Web data as academic and business quality estimates: A comparison of three data sources, J. Am. Soc. Inf. Sci. Technol. **63**(10), 1960–1972 (2012)
33.110 K. Kousha, M. Thelwall: Motivations for URL citations to open access library and information science articles, Scientometrics **68**(3), 501–517 (2006)
33.111 D. Stuart, M. Thelwall: Investigating triple helix relationships using URL citations: A case study of the UK West Midlands automobile industry, Res. Evaluation **15**(2), 97–106 (2006)
33.112 D. Wilkinson, M. Thelwall: Search markets and search results: The case of Bing, Libr. Inf. Sci. Res. **35**(4), 318–325 (2013)
33.113 M. Thelwall: Extracting accurate and complete results from search engines: Case study windows live, J. Am. Soc. Inf. Sci. Technol. **59**(1), 38–50 (2008)
33.114 P. Sud, M. Thelwall: Linked title mentions: A new automated link search candidate, Scientometrics **101**(3), 1831–1849 (2014)
33.115 F. Barjak, M. Thelwall: A statistical analysis of the web presences of European life sciences research teams, J. Am. Soc. Inf. Sci. Technol. **59**(4), 628–643 (2008)
33.116 F. Barjak, X. Li, M. Thelwall: Which factors explain the web impact of scientists personal homepages?, J. Am. Soc. Inf. Sci. Technol. **58**(2), 200–211 (2007)
33.117 A. Mas-Bleda, M. Thelwall, K. Kousha, I.F. Aguillo: Do highly cited researchers successfully use the social web?, Scientometrics **101**(1), 337–356 (2014)
33.118 A. Mas-Bleda, M. Thelwall, K. Kousha, I. Aguillo: Successful researchers publicizing research online: An outlink analysis of European highly cited scientists personal websites, J. Doc. **70**(1), 148–172 (2014)
33.119 S. Meece, A. Robinson, M.T. Gramstadt: Engaging researchers with the world's first scholarly arts repositories: Ten years after the UK's Kultur project, New Rev. Acad. Librariansh. **23**(2/3), 209–232 (2017)
33.120 M.S. Mayernik, D.L. Hart, K.E. Maull, N.M. Weber: Assessing and tracing the outcomes and impact of research infrastructures, J. Assoc. Inf. Sci. Technol. **68**(6), 1341–1359 (2017)
33.121 K.E. Eccles, M. Thelwall, E.T. Meyer: Measuring the web impact of digitised scholarly resources, J. Doc. **68**(4), 512–526 (2012)
33.122 C. Warwick, M. Terras, P. Huntington, N. Pappa: If you build it will they come? The LAIRAH study: Quantifying the use of online resources in the arts and humanities through statistical analysis of user log data, Lit. Linguist. Comput. **23**(1), 85–102 (2008)
33.123 A. Zuccala, M. Thelwall, C. Oppenheim, R. Dhiensa: Web intelligence analyses of digital libraries: A case study of the national electronic library for health (NeLH), J. Doc. **63**(4), 558–589 (2007)
33.124 D. Duin, D. King, P. Van Den Besselaar: Identifying audiences of e-infrastructures-tools for measuring impact, PLoS One **7**(12), e50943 (2012)
33.125 K. Jonkers, G.E. Derrick, C. Lopez-Illescas, P. Van den Besselaar: Measuring the scientific impact of e-research infrastructures: A citation based approach?, Scientometrics **101**(2), 1179–1194 (2014)
33.126 The London School of Economics and Political Science: LSE impact blog, http://blogs.lse.ac.uk/impactofsocialsciences/
33.127 DC's Improbable Science: Why you should ignore altmetrics and other bibliometric nightmares, http://www.dcscience.net/2014/01/16/why-you-should-ignore-altmetrics-and-other-bibliometric-nightmares/ (2014)
33.128 H. Shema, J. Bar-Ilan, M. Thelwall: Do blog citations correlate with a higher number of future citations? Research blogs as a potential source for alternative metrics, J. Assoc. Inf. Sci. Technol. **65**(5), 1018–1027 (2014)
33.129 H.R. Jamali, D. Alimohammadi: Blog citations as indicators of the societal impact of research: Content analysis of social sciences blogs, Int. J. Knowl. Content Dev. Technol. **5**(1), 15–32 (2015)
33.130 S. Kjellberg: Researchers' blogging practices in two epistemic cultures: The scholarly blog as a situated genre, Hum. IT **12**(3), 36–77 (2014)
33.131 H. Shema, J. Bar-Ilan, M. Thelwall: Research Blogs and the discussion of scholarly information, PLoS One **7**(5), e35869 (2012)
33.132 H. Shema, J. Bar-Ilan, M. Thelwall: How is research blogged? A content analysis approach, J. Assoc. Inf. Sci. Technol. **66**(6), 1136–1149 (2015)
33.133 J.K. Hoang, J. McCall, A.F. Dixon, R.T. Fitzgerald, F. Gaillard: Using social media to share your radiology research: How effective is a blog post?, J. Am. Coll. Radiol. **12**(7), 760–765 (2015)
33.134 N. Hookway, H. Snee: The Blogosphere. In: *The SAGE Handbook of Online Research Methods*, ed. by N.G. Fielding, R.M. Lee, G. Blank (SAGE, Oxford 2017) pp. 380–398
33.135 A.T. Weinstein, D.A. McFarlane: Case study–how a business school blog can build stakeholder relationships and create added value in an MBA marketing program, J. Strateg. Mark. **25**(2), 1–13 (2017)
33.136 Research Excellence Framework: Impact case study "China economic policy", http://impact.ref.ac.uk/CaseStudies/CaseStudy.aspx?Id=40126 (2014)

33.137 Research Excellence Framework: Impact case study "Stimulating public engagement with physics through the LHC and discovery of the Higgs boson", http://impact.ref.ac.uk/CaseStudies/CaseStudy.aspx?Id=34170 (2014)
33.138 V. Alberani, P.D.C. Pietrangeli, A.M. Mazza: The use of grey literature in health sciences: A preliminary survey, Bull. Med. Libr. Assoc. **78**(4), 358–363 (1990)
33.139 E. Orduna-Malea, A. Martin-Martin, E.D. Lopez-Cozar: Google Scholar and the gray literature: A reply to Bonato's review, ArXiv Preprint ArXiv:1702.03991
33.140 N.R. Haddaway, A.M. Collins, D. Coughlin, S. Kirk: The role of Google Scholar in evidence reviews and its applicability to grey literature searching, PLoS One **10**(9), e0138237 (2015)
33.141 S.E. Hug, M. Ochsner, M.P. Brändle: Citation analysis with Microsoft Academic, Scientometrics **111**(1), 371–378 (2017)
33.142 N.L. Pelzer, W.H. Wiese: Bibliometric study of grey literature in core veterinary medical journals, J. Med. Libr. Assoc. **91**(4), 434 (2003)
33.143 C. Talbot, C. Talbot: *Sir Humphrey and the Professors: What Does Whitehall Want from Academics?* (Univ. Manchester, Manchester 2014)
33.144 R. Cordes: Is grey literature ever used? Using citation analysis to measure the impact of GESAMP, an international marine scientific advisory body, Can. J. Inf. Libr. Sci. **28**(1), 49–69 (2013)
33.145 Q. Mahood, D.V. Eerd, E. Irvin: Searching for grey literature for systematic reviews: Challenges and benefits, Res. Synt. Methods **5**(3), 221–234 (2014)
33.146 A.M. Kalamar, S. Lee-Rife, M.J. Hindin: Interventions to prevent child marriage among young people in low-and middle-income countries: A systematic review of the published and gray literature, J. Adolesc. Health **59**(3), S16–S21 (2016)
33.147 G. McKiernan: ArXiv.org: The Los Alamos National Laboratory e-print server, Int. J. Grey Lit. **1**(3), 127–138 (2000)
33.148 X. Li, M. Thelwall, K. Kousha: The role of arXiv, RePEc, SSRN and PMC in formal scholarly communication, Aslib J. Inf. Manag. **67**(6), 614–635 (2015)
33.149 C. Seiler, K. Wohlrabe: Ranking economists on the basis of many indicators: An alternative approach using RePEc data, J. Informetr. **6**(3), 389–402 (2012)
33.150 C. Zimmermann: Academic rankings with RePEc, Econometrics **1**(3), 249–280 (2013)
33.151 Z. Kakushadze: An index for SSRN downloads, J. Informetr. **10**(1), 9–28 (2016)
33.152 D. Wilkinson, P. Sud, M. Thelwall: Substance without citation: Evaluating the online impact of grey literature, Scientometrics **98**(2), 797–806 (2014)
33.153 S.K. Panda: Shodhganga–a national level open access ETD repository of Indian electronic theses: Current status and discussions, Libr. Hi Tech News **33**(1), 23–26 (2016)
33.154 G. Marchionini, P. Solomon, C. Davis, T. Russell: Information and library science MPACT: A preliminary analysis, Libr. Inf. Sci. Res. **28**(4), 480–500 (2007)
33.155 C.R. Sugimoto, T.G. Russell, L.I. Meho, G. Marchionini: MPACT and citation impact: Two sides of the same scholarly coin?, Libr. Inf. Sci. Res. **30**(4), 273–281 (2008)
33.156 J. Kayongo, C. Helm: Relevance of library collections for graduate student research: A citation analysis study of doctoral dissertations at Notre Dame, Coll. Res. Libr. **73**(1), 47–67 (2011)
33.157 Y. Tonta, U. Al: Scatter and obsolescence of journals cited in theses and dissertations of librarianship, Libr. Inf. Sci. Res. **28**(2), 281–296 (2006)
33.158 M. Kettler: Ways of disseminating, tracking usage and impact of electronic theses and dissertations (ETDs), Grey J. TGJ **13**(1), 27–31 (2017)
33.159 T. Dewett, A. Denisi: Exploring scholarly reputation: It's more than just productivity, Scientometrics **60**(2), 249–272 (2004)
33.160 S. Jeong, S. Lee, H.G. Kim: Are you an invited speaker? A bibliometric analysis of elite groups for scholarly events in bioinformatics, J. Am. Soc. Inf. Sci. Technol. **60**(6), 1118–1131 (2009)
33.161 J. Yoo, S. Choi, M. Choi, J. Rho: Why people use Twitter: Social conformity and social value perspectives, Online Inf. Rev. **38**(2), 265–283 (2014)
33.162 M.A. Camilleri: Utilising content marketing metrics and social networks for academic visibility: Content marketing for academic impact. In: *Digital Tools for Academic Branding and Self-Promotion*, ed. by M. Cabrera, N. Lloret (IGI Global, Hershey 2016) pp. 109–126
33.163 M. Thelwall, K. Kousha: ResearchGate articles: Age, discipline, audience size and impact, J. Assoc. Inf. Sci. Technol. **68**(2), 468–479 (2017)
33.164 N. Busom, R. Petrlic, F. Sebé, C. Sorge, M. Valls: A privacy-preserving reputation system with user rewards, J. Netw. Comput. Appl. **80**(1), 58–66 (2017)
33.165 K. Jordan: *Understanding the structure and role of academics ego-networks on social networking sites* (The Open Univ, Milton Keynes 2017), Ph.D. Thesis
33.166 E. Orduna-Malea, A. Martín-Martín, M. Thelwall, E.D. López-Cózar: Do ResearchGate scores create ghost academic reputations?, Scientometrics **112**(1), 443–460 (2017)
33.167 S. Copiello, P. Bonifaci: A few remarks on ResearchGate score and academic reputation, Scientometrics (2018), https://doi.org/10.1007/s11192-017-2582-9
33.168 J. Poza-Lugan, A. Calduch-Losa: Social media as a tool, and tools offered by to social media to teachers and researchers: Personal or professional use? In: *Digital Tools for Academic Branding and Self-Promotion*, ed. by M. Cabrera, N. Lloret (IGI Global, Hershey 2016) pp. 66–80
33.169 M. Thelwall, K. Kousha: Academia.edu: Social network or academic network?, J. Assoc. Inf. Sci. Technol. **65**(4), 721–731 (2014)
33.170 M. Thelwall, K. Kousha: ResearchGate: Disseminating, communicating and measuring scholar-

ship?, J. Assoc. Inf. Sci. Technol. **66**(5), 876–889 (2015)
33.171 F. Riquelme, P. González-Cantergiani: Measuring user influence on Twitter: A survey, Inf. Process. Manag. **52**(5), 949–975 (2016)
33.172 M. Bouguessa, L.B. Romdhane: Identifying authorities in online communities, ACM Trans. Intell. Syst. Technol. (TIST) **6**(3), 30 (2015)
33.173 B. Cronin, H.W. Snyder, H. Rosenbaum, A. Martinson, E. Callahan: Invoked on the web, J. Am. Soc. Inf. Sci. **49**(14), 1319–1328 (1998)
33.174 C. Donovan, L. Butler: Testing novel quantitative indicators of research 'quality', esteem and 'user engagement': An economics pilot study, Res. Evaluation **16**(4), 231–242 (2007)
33.175 M.F. Fox: Women and men faculty in academic science and engineering: Social-organizational indicators and implications, Am. Behav. Sci. **53**(7), 997–1012 (2010)
33.176 K.E. Smith, E. Stewart: We need to talk about impact: Why social policy academics need to engage with the UK's research impact agenda, J. Soc. Policy **46**(1), 109–127 (2017)

Mike Thelwall

Faculty of Science and Engineering
University of Wolverhampton
Wolverhampton, UK
m.thelwall@wlv.ac.uk

Mike Thelwall, Professor of Data Science, leads the Statistical Cybermetrics Research Group at the University of Wolverhampton, UK, which he joined in 1989. He is also Docent at the Department of Information Studies at Åbo Akademi University and a research associate at the Oxford Internet Institute. His received his PhD was in Pure Mathematics from the University of Lancaster. Mike has developed tools and methods. He is an Associate Editor of the Journal of the Association for Information Science and Technology and sits on the Editorial Boards of the Journal of Informetrics, Scientometrics, and the Journal of Information Science. He was a member of the UK's independent review of the role of metrics in research assessment (2014–15) and is a member of the UK Forum for Responsible Research Metrics.

Part E

Advancement

Part E Advancement of Methodology for Patent Analysis

34. Information Technology-Based Patent Retrieval Models

Carson Leung, Wookey Lee, Justin Jongsu Song

This chapter presents information technology (IT) based patent retrieval models. It first compares and contrasts information retrieval (IR) with patent retrieval, and highlights their key differences. For instance, IR can be considered as a precision-oriented retrieval, whereas patent retrieval can be considered as a recall-oriented retrieval. The chapter then describes the boolean retrieval model, which was designed for IR but can be used for patent retrieval. To facilitate effective patent retrieval, a basic patent retrieval model is presented. With this model, representative keyword terms are extracted from the user query and are ranked according to their importance so that top-*k* relevant patents can be retrieved with irrelevant patents eliminated. Moreover, the chapter also presents some enhancements and extensions to the basic patent retrieval model, which include incorporation of relevance feedback, estimation of the importance of keyword terms, text preprocessing of patent documents, and handling of patent category frequency. In addition, two dynamic patent retrieval models are also described. These two models perform interactive patent retrieval via dispersion or accumulation to dynamically rank the patents. Experimental results with real-life datasets show that the models presented in this chapter outperformed many conventional search systems with respect to time and cost. While this chapter focuses on the theoretical aspects of IT based patent retrieval models which are of interest to IT specialists, practical illustrative examples in the chapter demonstrate the empirical aspects of patent retrieval models which are helpful to IT practitioners.

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_34

34.1 Patent Retrieval Versus Information Retrieval

With scientific and technological advances, *intellectual property* (IP) has attracted worldwide attention from both qualitative and quantitative points of view. Data are of such a large scale that practitioners face unprecedented challenges collect and analyze the IP information which is relevant to their daily tasks. As science and technology indicators, *patents* have played crucial roles to defend and protect the IP of inventors for their individual innovations and/or innovations in high-tech corporations, institutes, organizations and traditional industries. This is noticeably apparent through the tremendously increasing number of patent activities in recent years. For example, according to the United States Patent and Trademark Office (USPTO) [34.1], more than 576 000 patents were filed in calendar year 2012, which was approximately 160% greater than the number in 2002. The number further increased to near 630 000 filed patents in calendar year 2015, which included more than 589 000 utility patent applications (with US origin and foreign origin), more than 39 000 design patent applications, and more than 1000 plant patent applications (Fig. 34.1). This number is expected to be higher for the current calendar year. Moreover, the number of patents filed to other offices—such as the Canadian Intellectual Property Office (CIPO), the European Patent Office (EPO), the Japan Patent Office (JPO), the Korean Intellectual Property Office (KIPO), and the State Intellectual Property Office (SIPO) of China—also keeps growing.

With this increasing number of patents stored in databases, it becomes almost infeasible for a patent examiner to identify all the relevant patents without utilizing a special methodological power. In addition, as the patent portfolio grows, knowledge discovery in patent databases also becomes imperative. Moreover, these patent databases are very useful in many aspects. For instance, they can be used for measurement of technological performance such as science and technology indicators for performance assessment in research and development (R&D). Hence, quantitative methods, statistics and indicators built upon patent data that are relevant for studies of R&D systems are in demand.

Currently, most of those who search patents do so by applying one of the following techniques:

- A self-classification code system for a domestic domain
- The International Patent Classification (IPC) rule [34.2].

However, these search techniques have severe limitations in obtaining desirable searching outputs. For instance, prevailing patent search engines usually include too many unrelated results. Consequently, with high volumes of unrelated results, patent experts have to spend lots of time on manually refining the results.

Generally, there are fundamental differences between patent search and general information retrieval (IR) [34.3]. In high-level abstract terms, general IR can be considered as a *precision-oriented retrieval*, which focuses more on precision than recall. In contrast, patent search can be considered as a *recall-oriented retrieval*, which focuses more on recall than precision.

Note that *precision* measures the fraction of retrieved instances that are relevant

$$\text{Precision} = \frac{\text{Relevant retrieved instances}}{\text{Retrieved instances}} . \tag{34.1}$$

Precision is sometimes known as *positive predictive value* (PPV), which is the fraction of true positives (TPs) among all *outcome positives* including TPs and

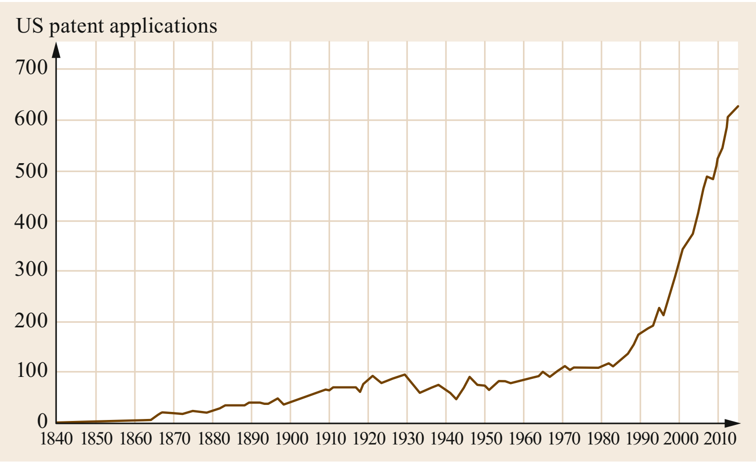


Fig. 34.1 Annual US patent statistics: Annual sum (in thousands) of all utility patent applications with US origin and foreign origin, design patent applications, and plant patent applications for the period 1840–2015 based on the USPTO statistics (after [34.1])

false positives (FPs)

$$\mathrm{PPV} = \frac{\mathrm{TPs}}{\mathrm{TPs} + \mathrm{FPs}} , \quad (34.2)$$

where outcome positives (i. e., TPs + FPs) are instances that are identified as *positive* outcomes regardless of whether they are correctly identified as positives (i. e., TPs) or incorrectly identified as positives (i. e., FPs). In other words, a FP is a *false alarm* or *Type-I error*. In the context of general retrieval:

1. Outcome positives are retrieved instances
2. TPs are relevant retrieved instances
3. FPs are irrelevant retrieved instances.

In contrast, *recall* measures the fraction of relevant instances that are retrieved

$$\mathrm{Recall} = \frac{\text{Relevant retrieved instances}}{\text{Relevant instances}} . \quad (34.3)$$

Recall is sometimes known as *sensitivity* or *true positive rate* (TPR), which is the fraction of TPs among all *condition positives* including TPs and false negatives (FNs)

$$\mathrm{TPR} = \frac{\mathrm{TPs}}{\mathrm{TPs} + \mathrm{FNs}} , \quad (34.4)$$

where condition positives (i. e., TPs + FNs) are instances with positive conditions regardless whether they are positive instances that are correctly identified (i. e., TPs) or positive instances that are incorrectly rejected/ignored (i. e., FNs). In other words, a FN is a *miss* or *Type-II error*. In the context of general retrieval:

1. Condition positives are relevant instances
2. TPs are retrieved relevant instances
3. FNs are missed/ignored relevant instances.

The purpose of a patent search is different from general IR. A conventional *general IR* from the news or the web aims to achieve high precision. For instance, when conducting a news search or a web search, users are keen on finding the answer to their query as quickly as possible and within the top-ranked list. Otherwise, they reformulate the query. In other words, users want to find one or a few relevant documents at the top of the ranked list that satisfy their need. In contrast, a *patent search*—or recall-oriented retrieval in general—aims to achieve high recall. For instance, when conducting a patent search, patent examiners need to work through the ranked list to identify all relevant patents before they stop and reformulate the query. Similarly, when conducting a legal search, lawyers or legal officers need to work through a ranked list to find every piece of evidence related to the case at hand from the documents that are under legal holds before they stop and reformulate the query. In similar situations, scientists do not want to miss any important prior works related to their ongoing research, so they examine the ranked list to find all relevant works before they stop and reformulate the query. In these three instances, users of recall-oriented retrieval such as patent search are professional experts in the field of search, and they want to find all possible relevant documents within a ranked list. In other words, users are keen on finding as many as possible, if not all of, the relevant documents that satisfy their needs. In real life, a patent examiner typically checks hundreds to thousands of documents in the result list to locate all possible relevant documents [34.4]. Finding relevant documents at the top of the list remains a desirable feature in recall-oriented retrieval, because this may reduce the retrieval effort. However, the key objective is to retrieve all the relevant documents while trying to minimize the number of documents to be checked.

The most frequently used measures for general IR search assessment are *precision* and *recall*, as defined in (34.1) and (34.3), respectively. In an extreme case, to achieve high precision (at a price of low recall), a system could retrieve a single relevant instance without attempting to retrieve some or all relevant instances. This leads to a precision value of 1 but a very low recall value, especially when the number of relevant instances is high. To elaborate, let $d_1, \ldots, d_n$ be n relevant instances. Without loss of generality, let d_1 be the only relevant instance retrieved by the system. Then, the precision value of this search is $1/1 = 1$ according to (34.1), but the corresponding recall value according to (34.3) is $1/n$ which becomes very low for a high number n of relevant instances. Hence, such a high-precision (but low-recall) system is suitable for real-life applications in which users just want to retrieve a relevant instance satisfying the user queries.

However, there are many other real-life applications in which users want to retrieve as many relevant instances as possible that satisfy the user queries. An example of such applications is patent search, which calls for high-recall systems. In an extreme case, to achieve high recall (at a price of low precision), a system could retrieve all instances without attempting to distinguish which ones are relevant and which ones are not. This leads to a recall value of 1 but a very low precision, especially when the number of retrieved instances is high. To elaborate, let $d_1, \ldots, d_n$ be n instances retrieved by the system. Without loss of generality, let $d_1, \ldots, d_q$ be q relevant retrieved instances (where $q \leq n$). If the system retrieves all q relevant instances together with additional $(n - q)$ irrelevant instances, then the recall value of this search is $q/q = 1$ according to (34.3), but the corresponding precision value according to (34.1) is q/n which becomes very low for a high number n (where $q \ll n$) of retrieved instances. A potential problem of using such a high-recall (but low-precision) system is that it may retrieve numerous instances, out of which many are irrelevant. These irrelevant retrieved instances are considered as *noise*. The noise may force users to spend lots of time on refining the search or on finding relevant instances from a haystack of irrelevant instances.

To assess a retrieval system that does not go to these extreme cases and maintains a good balance of both precision and recall, an *F-score* (F_1) can be used for a fixed number of retrieved documents [34.5]

$$F_1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} . \tag{34.5}$$

As the F_1 score is designed for classification tasks rather than recall-oriented retrieval in general (or patent search in particular), the same recall value—and thus the same F_1 score—may be obtained when retrieving n documents from two systems in which relevant documents are retrieved in one system earlier than another. For example, suppose there are 11 relevant instances. System 1 retrieves seven instances out of which the first three are relevant; system 2 also retrieves seven instances out of which the last three are relevant. Then, the recall value and F_1 score for system 1 are $3/11$ and

$$\frac{2 \times \frac{3}{7} \times \frac{3}{11}}{\frac{3}{7} + \frac{3}{11}} = \frac{1}{3}$$

respectively, which yield the same recall value and F_1 score for system 2.

As an enhancement, when assessing recall-oriented retrieval or patent search, more emphasis should be focused on the *recall* value. Hence, F-score can be modified to measure a weighted combination of *average precision* (AP) [34.6] and recall. The AP is the mean of the precision values after each relevant instance is retrieved

$$AP = \frac{\sum_k P(k)}{\text{Relevant retrieved instances}} , \tag{34.6}$$

where $P(k)$ is the precision for *k relevant instances*. It measures the precision at r, denoted as $P@r$, where r is the rank of k-th relevant retrieved instance: $P@r = k/r$, which indicates that k of the r retrieved instances are relevant. In other words, $P(k)$ captures the minimum number of instances that need to be retrieved in order to obtain k relevant instances

$$P(k) = \frac{k \text{ relevant retrieved instances}}{\min \#\text{retrieved instances for } k \text{ relevant ones}} . \tag{34.7}$$

For example, suppose system 1 retrieves seven instances, out of which the first three are relevant. Then, the AP for system 1 is

$$\frac{\frac{1}{1} + \frac{2}{2} + \frac{3}{3}}{3} = 1 ,$$

because the first relevant instance is obtained after retrieving one instance, the first two relevant instances are obtained after retrieving two instances, and the first three relevant instances are obtained after retrieving three instances. In contrast, suppose system 2 also retrieves seven instances, out of which the last three are relevant. Then, the AP for system 2 is lower than that for system 1. Specifically, the AP for system 2 is

$$\frac{\frac{1}{5} + \frac{2}{6} + \frac{3}{7}}{3} = \frac{101}{315} < \frac{1}{3} ,$$

because the first relevant instance is obtained after retrieving five instances, the first two relevant instances are obtained after retrieving six instances, and the first three relevant instances are obtained after retrieving seven instances.

With the definition of AP, the computation of F-score can be modified as follows

$$F'_\beta = \frac{(1+\beta) \times \text{AP} \times \text{recall}}{\beta \times \text{AP} + \text{recall}} , \tag{34.8}$$

where β is the weight of recall to precision. With the computation of F'_β score, two systems that retrieve the same relevant documents would give the same recall value and F_1 score. However, the system that retrieves relevant documents earlier would give a higher F'_β score (where $\beta > 1$) than the one that retrieves relevant documents later. For example, suppose there are 11 relevant instances. System 1 retrieves seven instances out of which the first three are relevant; system 2 also retrieves seven instances out of which the last three are relevant. Then, the recall value and F_1 score for system 1 are $3/11$ and

$$\frac{2 \times \frac{3}{7} \times \frac{3}{11}}{\frac{3}{7} + \frac{3}{11}} = \frac{1}{3}$$

respectively, which are the same recall value and F_1 score for system 2. However, the AP for system 1 is

$$\frac{\frac{1}{1} + \frac{2}{2} + \frac{3}{3}}{3} = 1 ,$$

which is higher than the AP for system 2 with a value of

$$\frac{\frac{1}{5} + \frac{2}{6} + \frac{3}{7}}{3} = \frac{101}{315} < \frac{1}{3} .$$

Thus, the F'_β score for system 1 is higher than that for system 2 (where $\beta > 1$).

34.2 Boolean Retrieval Model

Patent retrieval is one of the main activities carried out in patent offices. It checks the novelty of patent applications for the filed inventions. For many years, the boolean search has been a common approach used for finding relevant patents and documents that can invalidate the novelty of a patent application. Although this approach is very exhaustive and time consuming, interactive boolean search remains the search technique preferred by patent examiners because it is reproducible [34.4]. Here, *reproducibility* of patent retrieval means that the retrieval system will always give the same results for the same query each time. Having reproducible retrieval results is an essential criterion for patent examiners to defend their decisions about the novelty of patents.

To elaborate, as the boolean search is reproducible, it gives the same results for the same query—even when irrelevant documents are added to the document collection. Of course, the search may give a different result when relevant documents are added to the document collection. Note that a reproducible search is different from a probabilistic search, which may give a different result even when irrelevant documents are added to the document collection.

34.2.1 Extended Boolean Retrieval Model

Many existing patent retrieval systems originate from the boolean model [34.4, 7–9], which returns all the patent documents if any search keyword in the query is included. The query can be extended by applying boolean operators (e. g., AND, OR, NOT) on the keywords in the query. As an addition to the patent retrieval systems like USPTO, several quantifiers—such as AND, OR, NEAR, and NOT—are added to form an *extended boolean model*. On the one hand, the search results are reduced when the query is restrained with AND operators (i. e., when the number of AND operators is increased in the query). On the other hand, the search results are increased when the query is relaxed with OR operators (i. e., when the number of OR operators is increased in the query).

Moreover, although both the boolean retrieval model and its extended boolean retrieval model are commonly used for both IR and patent retrieval due to their reproducibility, they can be time-consuming. For precision-oriented tasks like IR, as the focus is on retrieving the top few relevant results, the high runtime complexity may not be a big concern. However, it becomes a big concern for recall-oriented tasks like patent searches because the focus is on retrieving all relevant results. Hence, it is desirable to have a model designed for (recall-oriented) patent retrieval, which focuses on retrieving all relevant patents.

34.3 Basic Patent Retrieval Model

When the number of searched patents increases, the amount of both valid data (i. e., relevant patents) and noise data (i. e., irrelevant patents) examined by the boolean retrieval model also increases. Consequently, it may take a lot of time to eliminate the noise from the valid candidate set. Knowing that the number of patents to be searched can be reduced once the recall is satisfied, the amount of noise data can be eliminated, and thus the possibility of losing valid data can also be reduced. So, noise elimination has become important for any patent retrieval model [34.10]. This section describes key phases of a basic patent retrieval model that finds the top-k valid answers (i. e., relevant patents) and eliminates noise (i. e., irrelevant patents).

34.3.1 Extraction of Representative Terms

As patent search queries consist of many keywords (e. g., hundreds of terms) and need thorough investigation of the results, a patent search usually takes much more time than a conventional general IR search from the web. A key success factor for a patent search relies on the selection of good search keywords. The keyword selection phase identifies the importance of the terms on which a proper weight needs to be assigned. One way to extract representative keyword terms is to model the term distribution of a patent query. By doing so, a detailed representation of the patent query can be obtained.

34.3.2 Ranking of Extracted Terms Based on Term Frequency

After extracting the representative keyword terms from the user query and a collection D of patent documents, the next phase is to rank these terms according to their importance. In the context of patent search, the quality of a search is influenced by the distribution of keyword terms. Note that there are two kinds of distribution for keyword terms:

- Absolute distribution
- Relative distribution.

The *term frequency* (tf) of a keyword t_i in a patent document d_j measures the absolute or relative word count $\mathrm{tf}(t_i, d_j)$ of t_i in d_j. It can be obtained by using keyword term measurements like the *term frequency-inverse document frequency* (TF-IDF). More specifically, the *absolute* term frequency of a keyword t_i in a patent document d_j measures the word count $\mathrm{tf}(t_i, d_j)$ of t_i in d_j, whereas the *relative* term frequency of a keyword t_i in a document d_j measures the ratio of the word count of t_i in d_j to the total number of words in d_j, i. e.,

$$\frac{\mathrm{tf}(t_i, d_j)}{\sum_{t_h} \mathrm{tf}(t_h, d_j)} .$$

Consider a set $D = \{d_1, d_2, \ldots, d_n\}$ of n patent documents. Any document $d_j = \{t_1, t_2, \ldots, t_m\}$ of these n documents contains m keyword terms (where $1 \le j \le n$). Then, the keyword weight distribution $\mathrm{KWD}(d_j)$ of a document d_j (where $1 \le j \le n$ documents in D), which is the distribution among the fully-qualified keywords from d_j, is defined as a product of the keyword weights $\mathrm{KW}(t_i, d_j)$ for all m terms in document d_j

$$\mathrm{KWD}(d_j) = \prod_{i=1}^{m} \mathrm{KW}(t_i, d_j) , \tag{34.9}$$

where $1 \le i \le m$ terms in d_j. Here, the keyword weight $\mathrm{KW}(t_i, d_j)$—which is also known as *cardinality*—of a term t_i contained in a document d_j can be computed as follows

$$\mathrm{KW}(t_i, d_j) = \frac{1}{1 + \mathrm{e}^{-\frac{\mathrm{tf}(t_i, d_j)}{10}}} , \tag{34.10}$$

such that (a) $1 \le i \le m$ terms in d_j and (b) $1 \le j \le n$ documents in D. Note that, in order not to weight too heavily on the term frequency of keyword term t_i, a sigmoid function $f(x) = 1/(1 + \mathrm{e}^{-x})$ can be used (where $x \ge 0$).

Based on the definition of keyword weight $\mathrm{KW}(t_i, d_j)$ in (34.10), when keyword term t_i appears very rarely in document d_j (i. e., a very low $\mathrm{tf}(t_i, d_j)$), the corresponding keyword weight $\mathrm{KW}(t_i, d_j)$ approaches $1/(1 + 1) = 0.5$. On the other hand, when keyword term t_i appears very frequently in d_j (i. e., a very high $\mathrm{tf}(t_i, d_j)$), the corresponding keyword weight $\mathrm{KW}(t_i, d_j)$ approaches $1/(1 + 0) = 1$. Hence, the range of $\mathrm{KW}(t_i, d_j)$ is in the range of $[0.5, 1]$.

Then, since keyword weight distribution $\mathrm{KWD}(d_j)$ is defined in (34.9) as the product of $\mathrm{KW}(t_i, d_j)$ over all m terms in a document d_j, the value of $\mathrm{KWD}(d_j)$ can range from $(0.5)^m$ to 1^m. When keyword term t_i appears very rarely in d_j (i. e., a very low $\mathrm{tf}(t_i, d_j)$) for many of the m terms (where $1 \le i \le m$), the corresponding $\mathrm{KWD}(d_j)$ approaches $(0.5)^m$, which approximately converges to 0 for a very large number m of keyword terms. On the other hand, when keyword term t_i appears very frequently in d_j (i. e., a very high $\mathrm{tf}(t_i, d_j)$) for many of the m keyword terms, the corresponding $\mathrm{KWD}(d_j)$ approaches $1^m = 1$. Hence, the range of $\mathrm{KWD}(d_j)$ is approximately $(0, 1]$.

34.3.3 Retrieval of Top-k Answers with Elimination of Noise by the Patent Threshold Algorithm (Patent TA)

After extracting the representative keyword terms and ranking these extracted terms based on their keyword weight distribution (which shows the importance of the keyword terms), the next phase is to retrieve the top-k answers (i. e., relevant instances) to user query q while eliminating noise (i. e., irrelevant instances). Specifically, for m keyword terms in n documents, a total of m lists are created. A list L_i of length n is created for each keyword t_i for $1 \le i \le m$. The j-th entry (where $1 \le j \le n$) in the list L_i captures the keyword weight $\mathrm{KW}(t_i, d_j)$ of t_i in document d_j. Entries in each list are then sorted in descending order of their keyword weights.

The patent retrieval process starts with the top entry in each of these m sorted lists (i. e., the highest keyword weight of each t_i for $1 \le i \le m$). Set a new threshold τ for the patent retrieval and noise elimination process to be the lowest keyword weight among the m keyword weights at the tops of these lists. For each document d_y appearing at the top of these lists find its lowest keyword weight within d_y (i. e., $\min_{1 \le i \le m} \mathrm{KW}(t_i, d_y)$). If such a lowest keyword weight in d_y meets or exceeds τ (i. e., $\min_{1 \le i \le m} \mathrm{KW}(t_i, d_y) \ge \tau$) implying that all m keyword weights within d_y meet or exceed τ (i. e., $\mathrm{KW}(t_i, d_y) \ge \tau$ for all $i \in [1, m]$), then d_y is chosen as one of the top-k answers to user query q. Otherwise, keep such a lowest keyword weight in d_y for future consideration, and continue with the next keyword weight on each of the m sorted lists.

Afterwards, set a new threshold τ for the patent retrieval and noise elimination process to be the lowest keyword weight among these m keyword weights (i. e., second highest keyword weights on these m lists). For each document d_y appearing at the new top of these lists, find its lowest keyword weight within d_y. If such a lowest keyword weight in d_y meets or exceeds τ, then d_y

is chosen as one of the top-k answers to user query q. Otherwise, repeat this process until all top-k answers are found and returned to the user. This patent retrieval process is called the *patent threshold algorithm* (*patent TA*). A benefit of using this patent TA is that it can find the top-k answers without examining the entire $m \times n$ list entries.

34.3.4 An Illustrative Example of the Basic Patent Retrieval Model

As an illustrative example, consider a practical situation where the user would like to search patents about service robots for vacuum cleaning. Table 34.1 shows the keyword term frequencies based on $m = 2$ keyword terms (`service robot`, `vacuum cleaning`) extracted from $n = 4$ sample documents (d_1, d_2, d_3 and d_4). Here, keyword term t_1 = `service robot` appears 80, 60, 30 and 40 times in d_1, d_2, d_3 and d_4, respectively—i. e., $\mathrm{tf}(t_1, d_1) = 80$, $\mathrm{tf}(t_1, d_2) = 60$, $\mathrm{tf}(t_1, d_3) = 30$ and $\mathrm{tf}(t_1, d_4) = 40$. Keyword term t_2 = `vacuum cleaning` appears 70, 50, 90 and 20 times in d_1, d_2, d_3 and d_4, respectively. Hence, keyword weights of these two terms in the four documents can be computed using (34.10). For instance,

$$\begin{aligned}
&\mathrm{KW}(t_1, d_1) = 0.9997\,, && \mathrm{KW}(t_1, d_2) = 0.9975\,,\\
&\mathrm{KW}(t_1, d_3) = 0.9526\,, && \mathrm{KW}(t_1, d_4) = 0.9820\,,\\
&\mathrm{KW}(t_2, d_1) = 0.9991\,, && \mathrm{KW}(t_2, d_2) = 0.9933\,,\\
&\mathrm{KW}(t_2, d_3) = 0.9999\,, \quad \text{and} && \mathrm{KW}(t_2, d_4) = 0.8808\,,
\end{aligned}$$

as shown in Table 34.2.

Table 34.1 Sample keyword term frequencies for a patent search on `service robots for vacuum cleaning` with the basic patent retrieval model

$\mathrm{tf}(t_i, d_j)$	t_1 = `service robot`	t_2 = `vacuum cleaning`
Document d_1	80	70
Document d_2	60	50
Document d_3	30	90
Document d_4	40	20

Table 34.2 Keyword weights for the sample keyword term frequencies in Table 34.1 for a patent search on `service robots for vacuum cleaning` with the basic patent retrieval model

$\mathrm{KW}(t_i, d_j)$	t_1 = `service robot`	t_2 = `vacuum cleaning`
Document d_1	0.9997	0.9991
Document d_2	0.9975	0.9933
Document d_3	0.9526	0.9999
Document d_4	0.9820	0.8808

To retrieve the top-2 answers (i. e., top-2 patent documents) with noise elimination, documents are ranked in descending order of their keyword weights. Specifically, the sorted list L_1 for keyword term t_1 contains $\langle d_1, 0.9997\rangle$, $\langle d_2, 0.9975\rangle$, $\langle d_4, 0.9820\rangle$, then $\langle d_3, 0.9526\rangle$. Similarly, the sorted list L_2 for keyword term t_2 contains $\langle d_3, 0.9999\rangle$, $\langle d_1, 0.9991\rangle$, $\langle d_2, 0.9933\rangle$, then $\langle d_4, 0.8808\rangle$. Set a threshold τ to be 0.9997, which is the minimum of $\mathrm{KW}(t_1, d_1) = 0.9997$ and $\mathrm{KW}(t_2, d_3) = 0.9999$. For documents (i. e., d_1 and d_3) appearing at the top of these two lists L_1 and L_2, their corresponding lowest keyword weights $\mathrm{KW}(t_2, d_1) = 0.9991$ and $\mathrm{KW}(t_1, d_3) = 0.9526$ are found. As neither of these two weights meets or exceeds $\tau = 0.9997$, they are kept for future consideration.

Then, continue the patent retrieval process with the next entry on each of these two sorted lists, i. e., $\langle d_2, 0.9975\rangle$ on L_1 and $\langle d_1, 0.9991\rangle$ on L_2. Update τ to be 0.9975, which is the minimum of these two weights. The lowest keyword weight in d_2 is $\mathrm{KW}(t_2, d_2) = 0.9933$. Recall that the lowest keyword weight in d_1 (i. e., $\mathrm{KW}(t_2, d_1) = 0.9991$) was found earlier. Consequently, the current top-2 lowest keyword weights become $\mathrm{KW}(t_2, d_1) = 0.9991$ and $\mathrm{KW}(t_2, d_2) = 0.9933$ because $\mathrm{KW}(t_1, d_3) = 0.9526$ possesses a lower weight than these top-2 lowest keyword weights. Note that, as $\mathrm{KW}(t_2, d_1) = 0.9991 \geq \tau = 0.9975$, d_1 is returned as one of the top-2 answers. On the other hand, as $\mathrm{KW}(t_2, d_2) = 0.9933$ does not meet or exceed τ, it is kept for future consideration.

In a similar fashion, the next entries on these two sorted lists are $\langle d_4, 0.9820\rangle$ on L_1 and $\langle d_2, 0.9933\rangle$ on L_2. Update τ to be 0.9820, which is the minimum of these two weights. The lowest keyword weight in d_4 is $\mathrm{KW}(t_2, d_4) = 0.8808$. Recall that the lowest keyword weight in d_2 (i. e., $\mathrm{KW}(t_2, d_2) = 0.9933$) was found earlier. Consequently, the current top-2 lowest keyword weights remain $\mathrm{KW}(t_2, d_1) = 0.9991$ and $\mathrm{KW}(t_2, d_2) = 0.9933$ because $\mathrm{KW}(t_2, d_4)$ possesses a lower weight than these top-2 lowest keyword weights. Note that, as $\mathrm{KW}(t_2, d_2) = 0.9933 \geq \tau = 0.9820$, d_2 is returned by the basic patent retrieval model. This completes the patent retrieval process with top-2 answers (namely, d_1 and d_2) returned to the user.

34.3.5 Summary

This basic model retrieves the top-k valid answers with noise elimination (i. e., retrieving top-k relevant patents with the elimination of irrelevant patents) by first extracting representative keyword terms from user patent-query q and the target patent collection D.

These terms are then ranked according to the keyword weight distribution (KWD), which is computed based on term frequency $\mathrm{tf}(t_i, d_j)$ of a keyword t_i in a patent document d_j. Afterwards, top-k answers to user query q can then be found by using the top-k patent threshold algorithm (patent TA), which can be considered as an adaptation of the threshold algorithm [34.11] designed for general IR. As patent TA was designed specifically for patent retrieval, it efficiently retrieves the top-k patents while eliminating noise (i. e., removing irrelevant instances).

34.4 Enhancements and Extensions to the Basic Patent Retrieval Model

While the basic model described in Sect. 34.3 retrieves the top-k patents with noise elimination, such a model can be further enhanced and extended. This section describes some of these enhancements and extensions.

34.4.1 Relevance Feedback

Recall from Sect. 34.3.1 that the first phase of the basic patent retrieval model is to extract representative keyword terms. Given a patent database (e. g., the USPTO patent database), an enhancement to the basic patent retrieval model is to take a sample, from which representative keyword terms (and appropriate features) are extracted to form an initial query vector $\boldsymbol{q}_{\text{init}}$. Ideally, such an initial query vector $\boldsymbol{q}_{\text{init}}$ should be formed in a way that maximizes the similarities of relevant documents while minimizing the similarities of irrelevant documents. Let D_{R} be a set of relevant documents, and let D_{I} be a set of irrelevant documents. Then, $\boldsymbol{q}_{\text{init}}$ should be formed as follows [34.12]

$$\boldsymbol{q}_{\text{init}} = \arg\max_{\boldsymbol{q}} \left(\mathrm{sim}(\boldsymbol{q}, D_{\text{R}}) - \mathrm{sim}(\boldsymbol{q}, D_{\text{I}})\right) , \quad (34.11)$$

where sim() can be any function that measures similarity. A commonly used similarity function is *cosine similarity*, under which the optimal query vector $\boldsymbol{q}_{\text{opt}}$ for separating the relevant and irrelevant patent documents can be derived as follows

$$\boldsymbol{q}_{\text{opt}} = \frac{\sum_{\boldsymbol{d}_j \in D_{\text{R}}} \boldsymbol{d}_j}{|D_{\text{R}}|} - \frac{\sum_{\boldsymbol{d}_j \in D_{\text{I}}} \boldsymbol{d}_j}{|D_{\text{I}}|} , \quad (34.12)$$

which means that the optimal query is the vector difference between the centroids of the relevant and irrelevant patent documents [34.13]. Unfortunately, complete knowledge of D_{R} and D_{I} may not always be available in practice.

Instead, a practical enhancement is to incorporate *relevance feedback* (e. g., some partial knowledge of known relevant and irrelevant documents). A well-known relevance feedback mechanism is the Rocchio model [34.14]. With this mechanism, a suboptimal query $\boldsymbol{q}_{\text{s}}$ can be computed as follows

$$\boldsymbol{q}_{\text{s}} = \alpha \boldsymbol{q}_{\text{init}} + \beta \frac{\sum_{\boldsymbol{d}_j \in D_{\text{KR}}} \boldsymbol{d}_j}{|D_{\text{KR}}|} + \gamma \frac{\sum_{\boldsymbol{d}_j \in D_{\text{KI}}} \boldsymbol{d}_j}{|D_{\text{KI}}|} , \quad (34.13)$$

where (a) D_{KR} and D_{KI} are sets of *known* relevant and irrelevant patent documents, respectively, whereas (b) α, β and γ are weights attached to each term.

34.4.2 Estimation of the Importance of Keyword Terms

In addition to incorporating relevance feedback, another enhancement is to bridge the vocabulary gap between the underlying information for a user patent-query q and the target patent collection D by setting up a unigram model of a query's language θ_q to *estimate the importance of each keyword term* t_i according to a weighted log-likelihood based approach [34.15]

$$P(t_i|q) = P(t_i|\theta_q) \log\left(\frac{P(t_i|\theta_q)}{P(t_i|\theta_D)}\right) . \quad (34.14)$$

Here, the maximum likelihood estimate $P(t_i|\theta_q)$ is a weight for a query term $t_i \in q$, which is defined as follows

$$P(t_i|\theta_q) = (1-\lambda) P_{\text{ML}}(t_i|d_j) + \lambda P_{\text{ML}}(t_i|D) , \quad (34.15)$$

such that:

- $0 \le \lambda \le 1$
- The maximum likelihood estimate $P_{\text{ML}}(t_i|d_j)$ of the query term t_i in a patent document d_j is calculated as follows

  $$P_{\text{ML}}(t_i|d_j) = \frac{\mathrm{tf}(t_i, d_j)}{\sum_{t_h} \mathrm{tf}(t_h, d_j)} , \quad (34.16)$$

  for $\mathrm{tf}(t_i, d_j)$ denotes the *term frequency* of t_i in d_j, which can be obtained by using keyword term measurements like the *term frequency-inverse document frequency* (*TF-IDF*)

- The maximum likelihood estimate $P_{\text{ML}}(t_i|D)$ of the query term t_i in a collection D of documents (where $D = \{d_j|1 \le j \le n\}$) is calculated as follows

$$P_{\text{ML}}(t_i|D) = \frac{\text{tf}(t_i, D)}{\sum_{t_h} \text{tf}(t_h, D)} . \quad (34.17)$$

Equation (34.14) captures the *Kullback–Leibler* divergence [34.15] between the document language model θ_{q} and the collection language model θ_D. Those terms with high similarity to θ_q and θ_D are favorable. Moreover, such an estimate $P(t_i|q)$ in the equation can be normalized by a normalization factor

$$\frac{1}{\sum_{t_i \in q} P(t_i|q)} .$$

34.4.3 Patent Text Preprocessing

Recall from Sect. 34.3.2 that the second phase of the basic patent retrieval model is to rank keyword terms in the patent documents. Hence, an extension to the basic patent retrieval model is to perform *text preprocessing* on these patent documents. The main purpose of this extension is for cleaning and preprocessing the patent documents for further analysis. Preprocessing involves the following tasks:

- *Syntax tagging*, which uses an English part-of-speech (POS) tagger (e. g., log-linear POS tagger [34.16]) to distinguish words and terms in sentences based on their syntactic or morphological features
- *Word stemming*, which reduces inflected (or derived) words to their word stem, base or root form (i. e., generally a written word form) so that only nouns (including singular or collective nouns, singular proper nouns, plural nouns, and plural proper nouns) and verbs (including verbs in base form, past tense, gerunds or present participles, past participles, as well as non-third person singular present and third person singular present verbs) are reserved
- *Stop-word elimination*, which removes irrelevant but extremely common words (e. g., a, on, which, zero).

34.4.4 Ranking of Terms Based on Category Frequency

In general, patent documents can be gathered from a variety of sources, such as the Canadian Intellectual Property Office (CIPO), the European Patent Office (EPO), the Japan Patent Office (JPO), the Korean Intellectual Property Office (KIPO), the State Intellectual Property Office (SIPO) of China, as well as the United States Patent and Trademark Office (USPTO). For instance, patent documents can be collected from target patent databases such as the USPTO database. The group-level labels of the International Patent Classification (IPC) codes are tagged for the basic attributes of the indexing vocabulary. For each of the group-level labels, a certain number of patent documents can be gathered for further analysis. Let $D = \{d_1, d_2, \ldots, d_n\}$ be the set of n patent documents (where n is the total number of documents in D). Let c_h be a group-level category code (which is the main group in the IPC hierarchy); let $C = \{c_1, c_2, \ldots, c_p\}$ be the set of group-level category codes within the IPC hierarchy, where p is the total number of category codes in the IPC hierarchy. Then, the set $D(c_h)$ of documents in a category code c_h includes all the documents belonging to category code c_h, i. e., $D(c_h) = \{d_j \in c_h\}$. For example, if category code c_1 includes document d_1, then $D(c_1) = \{d_1\}$. Similarly, if category code c_2 includes documents d_2 and d_3, then $D(c_2) = \{d_2, d_3\}$.

Recall from Sect. 34.3.2 that representative keyword terms extracted from the user query q and a collection D of patent documents are ranked based on the term frequency $\text{tf}(t_i, d_j)$ of a keyword t_i in a patent document d_j. Due to the variety of sources for patent documents, an extension to the basic patent retrieval model is the ability to rank patent documents based on category frequency—more specifically, based on the frequency of a keyword t_i in patent documents belonging to a group-level category code c_h within the IPC hierarchy. These frequencies of patent documents can be represented in a matrix form—specifically, in a (document $\times$ IPC category)-matrix $\mathbf{W}^{d \times c}$, which can be computed as a matrix product of a (document $\times$ term)-matrix $\mathbf{W}^{d \times t}$ with a (term $\times$ IPC category)-matrix $\mathbf{W}^{t \times c}$.

Step 1. Construction of a (Document × Term)-Matrix

To build a (document $\times$ term)-matrix $\mathbf{W}^{d \times t}$, let $T = \{t_1, t_2, \ldots, t_m\}$ be a set of distinctive keyword terms extracted from the patent documents, where (a) t_i is a keyword term in T and (b) m is the total number of keyword terms in T. Generally, an IPC category covers a certain number of documents, and each document consists of multiple terms. A set of terms contained in a document d_j is denoted as $T(d_j) = \{t_i|t_i \in d_j\}$. Moreover, let $\text{tf}_{j,i} = \text{tf}(t_i, d_j)$ represent the frequency of term t_i in document d_j. For example, if five terms t_1, t_1, t_2, t_4 and t_5 (i. e., with duplicated t_1) appear in document d_1, then $T(d_1) = \{t_1, t_2, t_4, t_5\}$ and $\text{tf}_{1,1} = 2$, $\text{tf}_{1,2} = 1$, $\text{tf}_{1,4} = 1$ and $\text{tf}_{1,5} = 1$.

Then, let $T(c_h) = \{T(d_j)|d_j \in c_h\}$ represent the set of terms belonging to the documents in IPC category code c_h. When using the traditional *vector space model* (VSM) built for retrieval, documents are represented as vectors of terms. The score of each attribute in the vector can be calculated by the following *term frequency-inverse document frequency* (*TF-IDF*) formula [34.17], which says that the weight factor TF-IDF can be calculated as a product of TF and IDF

$$\text{TF-IDF}(t_i, d_j, D) = \text{tf}(t_i, d_j) \times \text{idf}(t_i, D)\,, \tag{34.18}$$

where $\text{idf}(t_i, D)$ is commonly defined as follows

$$\text{idf}(t_i, D) = \log\left(\frac{n}{n_i}\right). \tag{34.19}$$

Here, (a) $n = |\{d_j \in D\}|$ is the total number of documents in D, and (b) the IDF weight $n_i = |\{d_j \in D | t_i \in d_j\}|$ is number of documents in D containing the term t_i. Thus, each weight factor $\text{TF-IDF}(t_j, d_j, D) = w_{j,i}^{d\times t}$ can be represented as an entry or element in a (document $\times$ term)-matrix $\mathbf{W}^{d\times t}$ with $n \times m$ entries capturing the weights of m distinctive keyword terms on n patent documents. In other words,

$$\mathbf{W}^{d\times t} = \left[w_{j,i}^{d\times t}\right]_{n\times m}\,, \tag{34.20}$$

where

$$w_{j,i}^{d\times t} = \text{tf}_{j,i} \times \log\left(\frac{n}{n_i}\right). \tag{34.21}$$

Step 2. Construction of a (IPC Category × Term)-Matrix Based on Patent Category Frequency

Inspired by the concept of *term frequency-inverse corpus frequency* (TF-ICF) [34.18] as a term weighting scheme for clustering dynamic data streams, the TF-IDF formula in (34.18) and (34.21) can be extended into the *term frequency-inverse patent category frequency* (TF-IPCF) for measuring the weight $w_{h,i}^{c\times t}$ of a term t_i of a category c_h as follows

$$w_{h,i}^{c\times t} = \frac{\sum_{d_j \in D(c_h)} \frac{\text{tf}_{j,i}}{\max_{t_l \in T(d_j)} \text{tf}_{j,l}}}{|D(c_h)|} \log\left(\frac{p}{p_i}\right), \tag{34.22}$$

where:

- $\max_{t_l \in T(d_j)} \text{tf}_{j,l}$ is the maximum frequency over all terms in document d_j, which is used to standardize the term frequency in a document so as to avoid the influence of varied document sizes
- $\sum_{d_j \in D(c_h)} \text{tf}_{j,i} / (\max_{t_l \in T(d_j)} \text{tf}_{j,l})$ measures the normalized term frequency in category c_h
- $|D(c_h)|$ is the number of documents in category c_h, which is used to standardize the TF value so as to avoid the bias caused by different numbers of documents contained in a category
- $\log(p/p_i)$ calculates the IPCF value for term t_i, where (a) p is the total number of categories and (b) p_i is the number of categories in which the percentage of documents containing term t_i that meets or exceeds a specific threshold. The setting of this threshold can be used to treat those infrequent terms as noise.

Advantages of using TF-IPCF in the extended patent retrieval model (when compared to TF-IDF used in the basic patent retrieval model) include the following:

- TF-IPCF reduces the bias caused by varied document length
- TF-IPCF reduces the bias due to the different numbers of documents contained in a category
- TF-IPCF remedies the bias caused when a term appears in almost every category (which may happen when a category contains many documents).

After calculating the weight $w_{h,i}^{c\times t}$ for every term t_i of a category c_h, the corresponding row vector $\boldsymbol{c}_h$ in the (IPC category $\times$ term)-matrix $\mathbf{W}^{c\times t}$ can be generated. In other words,

$$\boldsymbol{c}_h = \left[w_{h,1}^{c\times t}, w_{h,2}^{c\times t}, \ldots, w_{h,m}^{c\times t}\right]_{1\times m}\,, \tag{34.23}$$

which captures the weight $w_{h,i}^{c\times t}$ of each of the m keyword terms in $T(c_h)$ where $T(c_h) = \{T(d_j) | d_j \in c_h\}$. Here, the weight $w_{h,i}^{c\times t}$ can be calculated as per (34.22).

Consequently, an (IPC category $\times$ term)-matrix $\mathbf{W}^{c\times t}$ can be represented as a $p \times m$ matrix capturing the weights of m distinctive keyword terms on patent documents belonging to p patent categories, as follows

$$\mathbf{W}^{c\times t} = \left[w_{h,i}^{c\times t}\right]_{p\times m}\,, \tag{34.24}$$

$$= [\boldsymbol{c}_h]_{p\times 1}\,, \tag{34.25}$$

$$= \begin{bmatrix} \boldsymbol{c}_1 \\ \boldsymbol{c}_2 \\ \vdots \\ \boldsymbol{c}_p \end{bmatrix}, \tag{34.26}$$

$$= \begin{bmatrix} w_{1,1}^{c\times t} & w_{1,2}^{c\times t} & \cdots & w_{1,m}^{c\times t} \\ w_{2,1}^{c\times t} & w_{2,2}^{c\times t} & \cdots & w_{2,m}^{c\times t} \\ \vdots & \vdots & \ddots & \vdots \\ w_{p,1}^{c\times t} & w_{p,2}^{c\times t} & \cdots & w_{p,m}^{c\times t} \end{bmatrix}. \tag{34.27}$$

As a further enhancement, the size of this matrix can be reduced. Specifically, by using a filtering process to discover distinctive terms, the discrimination power of a term can be computed through a term distribution, which in turn is computed by using a term standard deviation among categories. During this filtering process, the terms with lower standard deviations—which indicate weaker discriminative capability among categories—are removed. Here, the standard deviation is used as the discrimination recognition threshold because it is a simple way to calculate the variation among attributes. The standard deviation $\rho(t_i)$ of a term t_i among categories can be calculated as follow

$$\rho(t_i) = \sqrt{\frac{1}{p}\sum_{h=1}^{p}\left(w_{h,i}^{c\times t} - \overline{w}_i^t\right)^2}\,, \tag{34.28}$$

where $\overline{w}_i^t$ is the average (i. e., the mean) of $w_{h,i}^{c\times t}$ over all p categories $c_1,\ldots,c_p$. A term t_i with a smaller standard deviation is observed to have weaker discriminative ability between categories, and vice versa. The distinctive term can be an index term in the indexing vocabulary. Through this filtering process, the (IPC category $\times$ term)-matrix can shrink in size.

Step 3. Construction of a (Document × IPC category)-Matrix

Recall that, in step 1, the (document $\times$ term)-matrix $\mathbf{W}^{d\times t} = [w_{j,i}^{d\times t}]_{n\times m}$ with $n\times m$ entries for capturing the weights of m distinct keyword terms on n patent documents was constructed with every weight value $w_{j,i}^{d\times t}$ calculated by the traditional TF-IDF formula in (34.21). Similarly, in Step 2, the (IPC category $\times$ term)-matrix $\mathbf{W}^{c\times t} = [w_{h,i}^{c\times t}]_{p\times m}$ with $p\times m$ entries for capturing the weights of m distinct keyword terms on patent documents belonging to p IPC patent categories was constructed with every weight value $w_{h,i}^{c\times t}$ calculated by the TF-IPCF formula in (34.22).

By taking the transpose of $\mathbf{W}^{c\times t}$, a (term $\times$ IPC category)-matrix $\mathbf{W}^{t\times c} = [w_{i,h}^{t\times c}]_{m\times p}$ with $m\times p$ entries for capturing the weights of m distinct keyword terms on patent documents belonging to p IPC patent categories can be obtained

$$\mathbf{W}^{t\times c} = (\mathbf{W}^{c\times t})^{\mathrm{T}}\,. \tag{34.29}$$

Consequently, in the current step (i. e., Step 3), a (document $\times$ IPC category)-matrix $\mathbf{W}^{d\times c} = [w_{j,h}^{d\times c}]_{n\times p}$ with $n\times p$ entries for capturing the weights of distinct keyword terms on n patent documents belonging to p IPC patent categories can be constructed by performing matrix multiplication on the matrices $\mathbf{W}^{d\times t}$ and $\mathbf{W}^{t\times c}$ as follows

$$\mathbf{W}^{d\times c} = \mathbf{W}^{d\times t}\times\mathbf{W}^{t\times c}\,, \tag{34.30}$$

$$= \mathbf{W}^{d\times t}\times(\mathbf{W}^{c\times t})^{\mathrm{T}}\,. \tag{34.31}$$

34.4.5 Extension of the Patent Threshold Algorithm for Handling IPC Category

Recall from Sect. 34.3.3 that the third phase of the basic patent retrieval model is to use the patent threshold algorithm (patent TA) to retrieve the top-k relevant patent documents based on the weight $w^{d\times t}$ of m distinct keyword terms on n patent documents. With the availability of the weight $w^{d\times c}$ of m distinct keyword terms on patent documents belonging to p IPC patent categories, the patent TA can be extended to the top-k relevant patent documents based on the weight $w^{d\times c}$.

To elaborate, the extended patent TA starts with the top element in each of these p sorted vectors (i. e., the highest weight of each c_h for $1\le h\le p$). Set a new threshold τ for the patent retrieval and noise elimination process to be the lowest weight among the p weights at the front of these vectors (i. e., the top row in the resulting matrix). For each document d_y appearing at the front of these vectors, find its lowest weight within d_j (i. e., $\min_{1\le h\le p} w_{j,h}^{d\times c}$). If such a lowest weight in d_j meets or exceeds τ (i. e., $\min_{1\le h\le p} w_{j,h}^{d\times c}\ge\tau$) implying that all p weights within d_j meet or exceed τ (i. e., $w_{j,h}^{d\times c}\ge\tau$ for all $h\in[1,p]$), then d_j is chosen as one of the top-k answers to user query q. Otherwise, keep such a lowest weight in d_j for future consideration, and continue with the next row of the matrix. Afterwards, set a new threshold τ for the patent retrieval and noise elimination process to be the lowest weight among these p weights (in this second row of the matrix). For each document d_j appearing in this row, find its lowest weight within d_j. If such a lowest weight in d_j meets or exceeds τ, then d_j is chosen as one of the top-k answers to user query q. Otherwise, repeat this process until all top-k answers are found and returned to the user. This patent retrieval process is called the *extended patent TA*.

Please note that retrieving the top-k answers with this extended patent TA is similar to doing so with the patent TA used in the basic model (described in Sect. 34.3.3). As such, a benefit of using this model for patent retrieval is that it can find top-k answers without examining the entire $n\times p$ matrix space. Moreover, since $\mathbf{W}^{d\times c}$ is in a matrix form, the number of vectors to be examined by this model during the patent retrieval process can be reduced by using IPC category-based *singular value decomposition* (SVD) which identifies the more relevant documents and categories.

34.4.6 An Illustrative Example of the Enhanced or Extended Patent Retrieval Model

As an illustrative example, consider a practical situation where the user would like to search patents about service robots for vacuum cleaning. Let $C = \{c_1, c_2\}$ be a set of group-level category codes within the International Patent Classification (IPC) hierarchy such that:

- c_1 covers keyword term

  $t_1 =$ `service robot`,
  $t_2 =$ `entertainment robot`,
  $t_3 =$ `industrial robot`,
  $t_4 =$ `juggling robot`, and
  $t_5 =$ `military robot`;

  whereas
- c_2 covers keyword term $t_6 =$ `vacuum cleaning` and other keyword terms for related household chores.

Consider the following matrix $\mathbf{W}^{d\times c}$ that captures the above keyword terms in $p = 2$ categories extracted from $n = 4$ documents

$$\mathbf{W}^{d\times c} = \begin{bmatrix} 0.9998 & 0.9997 \\ 0.9991 & 0.9975 \\ 0.9820 & 0.9999 \\ 0.9933 & 0.9526 \end{bmatrix}.$$

To retrieve the top-2 answers with noise elimination, documents are ranked in descending order of their weights for each category, resulting in the following two sorted vectors

$$\boldsymbol{L}_1 = \begin{bmatrix} d_1 \\ d_2 \\ d_4 \\ d_3 \end{bmatrix} \text{ and } \boldsymbol{L}_2 = \begin{bmatrix} d_3 \\ d_1 \\ d_2 \\ d_4 \end{bmatrix}.$$

Then, set a threshold τ to be 0.9998, which is the minimum of $w_{d_1,c_1} = 0.9998$ and $w_{d_3,c_2} = 0.9999$. For documents (i. e., d_1 and d_3) appearing at the top of these two vectors L_1 and L_2, their corresponding lowest weights $\min_h w_{d_1,c_h} = 0.9997$ and $\min_h w_{d_3,c_h} = 0.9820$ are found. As neither of these two weights meets or exceeds $\tau = 0.9998$, they are kept for future consideration.

Next, continue the patent retrieval process with the next element on each of these two sorted vectors (i. e., d_2 on $\boldsymbol{L}_1$ and d_1 on $\boldsymbol{L}_2$). Update τ to be 0.9991, which is the minimum of the two weights $w_{d_2,c_1} = 0.9991$ and $w_{d_1,c_2} = 0.9997$. The lowest weight in d_2 is $\min_h w_{d_2,c_h} = 0.9975$. Recall that the lowest weight in d_1 (i. e., $\min_h w_{d_1,c_h} = 0.9997$) was found earlier. Consequently, the current top-2 lowest weights become $\min_h w_{d_1,c_h} = 0.9997$ and $\min_h w_{d_2,c_h} = 0.9975$ because $\min_h w_{d_3,c_h} = 0.9820$ possesses a lower weight than these top-2 lowest weights. Note that, on the one hand, as $\min_h w_{d_1,c_h} = 0.9997 \geq \tau = 0.9991$, d_1 is returned as one of the top-2 answers. On the other hand, as $\min_h w_{d_2,c_h} = 0.9975$ does not meet or exceed τ, it is kept for future consideration.

In a similar fashion, the next elements on each of these two sorted vectors are d_4 on $\boldsymbol{L}_1$ and d_2 on $\boldsymbol{L}_2$. Update τ to be 0.9933, which is the minimum of the two weights $w_{d_4,c_1} = 0.9933$ and $w_{d_2,c_2} = 0.9975$. The lowest weight in d_4 is $\min_h w_{d_4,c_h} = 0.9526$. Recall that the lowest weight in d_2 (i. e., $\min_h w_{d_2,c_h} = 0.9975$) was found earlier. Consequently, the current top-2 lowest keyword weights remain $\min_h w_{d_1,c_h} = 0.9997$ and $\min_h w_{d_2,c_h} = 0.9975$ because $\min_h w_{d_4,c_h} = 0.9526$ possesses a lower weight than these top-2 lowest weights. Note that, as $\min_h w_{d_2,c_h} = 0.9975 \geq \tau = 0.9933$, d_2 is returned by the patent retrieval model. This completes the patent retrieval process with top-2 answers (namely, d_1 and d_2) returned to the user.

As illustrated in our examples, which are consistent with some empirical results [34.12], the patent retrieval model usually leads to higher recall and/or checks fewer irrelevant patent documents than traditional approaches such as the boolean model for retrieving patents from the USPTO patent database. To elaborate, the *recall for k retrieved documents*—denoted as $R(k)$, which measures the recall after retrieving k relevant documents—for the patent retrieval model keeps improving and $R(k)$ for traditional approaches increases only slightly (or remains the same) when k increases. As a result, the gap between the two keeps widening when k increases, showing the benefits of applying the patent retrieval model. Moreover, to achieve a certain recall target (say, an 80% recall), the boolean model patent retrieval described in Sect. 34.2 checked 1800 patent documents, but the patent retrieval model described in this section only checked 650 patent documents, when searching for 58 relevant patents.

34.4.7 Summary

By incorporating the enhancements and extensions described in this section, the resulting model can retrieve the top-k patents with noise elimination by first extracting appropriate features of a sample taken from the patent database to form an initial query vector, which is then modified after incorporating relevance feedback (e. g., partial knowledge of known relevant and irrelevant patent documents). Then, the importance of keyword terms from the resulting query vector can be estimated. Regarding the patent documents, they can be preprocessed by syntax tagging, word stemming, and stop-word elimination. Moreover, weights are assigned to keyword terms in documents belonging to different categories based on the term frequency-inverse document frequency (TF-IDF) formula (34.21) and term frequency-inverse patent category frequency (TF-IPCF) formula (34.22) for capturing the TF-IDF and the TF-IPCF. The resulting (document × IPC category)-matrix is formed by multiplying its corresponding (document × term)-matrix and the transpose of the (IPC category × term)-matrix (34.31). Documents in the vectors that form the resulting matrix are ranked in non-ascending order of their weights. The top-k answers to the user query can then be found by using the top-k extended patent TA.

34.5 Dynamic Patent Retrieval Models

Recall that, in both the basic patent retrieval model and its enhancements described in Sect. 34.3 and 34.4, documents were ranked in descending order of their weights and patent TA (or extended patent TA) was used to retrieve the top-k answers to a user query using this ranked list containing keyword terms. As alternatives to the patent TA or extended patent TA, two dynamic patent retrieval models [34.19] can be used to retrieve the top-k answers. These two models are:

- Dispersion for dynamic ranking (DDR)
- Accumulation for dynamic ranking (ADR)

34.5.1 Dispersion for Dynamic Ranking

The idea behind the *dispersion for dynamic ranking* (*DDR*) model [34.19] is that, in the process of retrieving the top-k patent documents, if the first patent document in the ranked list is irrelevant, the furthest document (say, d_j) will be checked in the next iteration. Here, various distance measures—such as cosine distance $\mathrm{dist}(d_x, d_y)$ between two patent documents d_x and d_y—can be used

$$\mathrm{dist}(d_x, d_y) = 1 - \mathrm{sim}(d_x, d_y)\ , \tag{34.32}$$

where $\mathrm{sim}(d_x, d_y)$ measures the cosine similarity between the vectors $\boldsymbol{d}_x$ and $\boldsymbol{d}_y$ representing two patent documents d_x and d_y, which is defined as follows

$$\mathrm{sim}(d_x, d_y) = \frac{\boldsymbol{d}_x \cdot \boldsymbol{d}_y}{||\boldsymbol{d}_x||_2\, ||\boldsymbol{d}_y||_2}\ . \tag{34.33}$$

Hence, by substituting (34.33) into (34.32), the resulting equation can be written as follows

$$\mathrm{dist}(d_x, d_y) = 1 - \frac{\boldsymbol{d}_x \cdot \boldsymbol{d}_y}{||\boldsymbol{d}_x||_2\, ||\boldsymbol{d}_y||_2}\ . \tag{34.34}$$

Let d_j be the document furthest from the first irrelevant patent document in the ranked list. If d_j happens to be relevant, then its most similar document (i. e., the nearest document) among the unchecked patents will be considered in the following iteration. The remaining iterations can be executed in a similar matter. In other words, for each iteration, if a currently checked patent document is relevant, then retrieve its nearest document (i. e., the most similar one). Otherwise, retrieve its furthest document (i. e., the least similar one from the current document). The DDR terminates when it finds all of the top-k answers.

34.5.2 An Illustrative Example of Dispersion for Dynamic Ranking

As an illustrative example, consider a ranked list of eight patent documents $\{d_1, d_2, \ldots, d_8\}$ such that documents d_1–d_6 are irrelevant and only d_7–d_8 are relevant. See Fig. 34.2 for the location of these documents in multi-dimensional vector space. To find the top-2 relevant patent documents, DDR starts with d_1, which is irrelevant. Then, DDR checks document d_2 (which is the furthest from d_1, i. e., $\arg\max_j \mathrm{dist}(d_j, d_1) = 2$) and finds that d_2 is also irrelevant. Afterwards, DDR checks and finds irrelevant document d_3 (which is the furthest from d_2) and its furthest (but irrelevant) document d_4. DDR checks document d_8 (which is the furthest from d_4, i. e., $\arg\max_j\, dist(d_j, d_4) = 8$) next and finds that d_8 turns out to be relevant. After returning d_8 as a top-1 relevant patent document, DDR checks document d_7 (which is the nearest to d_8) and finds that d_7 is also relevant. DDR then stops because it finds all top-2 relevant patent documents d_8 and d_7 (after checking four irrelevant documents d_1, d_2, d_3 and d_4).

34.5.3 Accumulation for Dynamic Ranking

The idea behind the *accumulation for dynamic ranking* (*ADR*) model [34.19] is very similar to that of the DDR, except that the vector accumulated values are used. More specifically, when finding the next patent document, ADR considers the document that is furthest from the cumulative mean value of the checked irrelevant documents (DDR considers the document that is furthest from the most recently checked irrelevant document). In

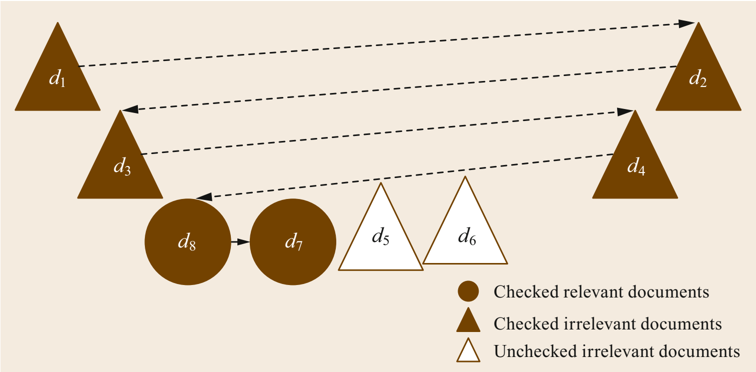


Fig. 34.2 Illustrative example on applying DDR to patent documents

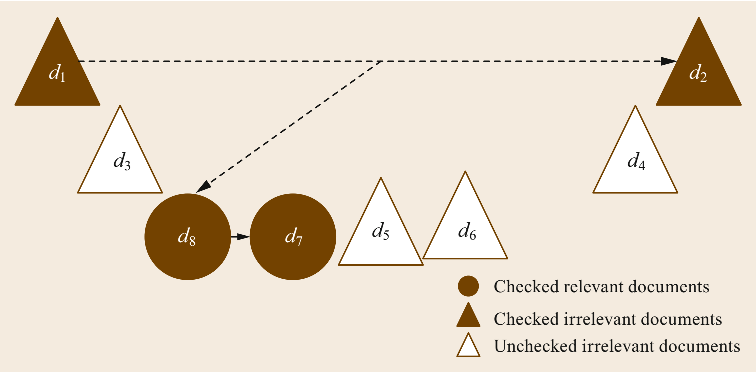


Fig. 34.3 Illustrative example on applying ADR to patent documents

other words, for each iteration, if a currently checked patent document is relevant, then retrieve its nearest document (i. e., the most similar one). Otherwise, retrieve the document that is furthest from the *vector average* among all previously checked irrelevant documents. The ADR terminates when it finds all top-k answers.

34.5.4 An Illustrative Example of Accumulation for Dynamic Ranking

As an illustrative example, consider a ranked list of eight patent documents $\{d_1, d_2, \ldots, d_8\}$ such that documents d_1–d_6 are irrelevant and only d_7–d_8 are relevant. See Fig. 34.3 for the location of these documents in multi-dimensional vector space. To find the top-2 relevant patent documents, ADR starts with d_1, which is irrelevant. Then, ADR checks document d_2 (which is the furthest from d_1, i. e., $\arg\max_j \mathrm{dist}(d_j, \mathrm{avg}(\{d_1\})) = 2$) and finds that d_2 is also irrelevant. Afterwards, ADR checks document d_8 (which is the furthest from the cumulative mean between d_1 and d_2, i. e., $\arg\max_j \mathrm{dist}(d_j, \mathrm{avg}(\{d_1, d_2\})) = 8$) and finds that d_8 turns out to be relevant. After returning d_8 as a top-1 relevant patent document, ADR checks document d_7 (which is the nearest to d_8) and finds that d_7 is also relevant. ADR then stops because it finds all top-2 relevant patent documents d_8 and d_7 (after checking only two irrelevant documents d_1 and d_2).

34.5.5 Summary

This section describes both the dispersion for dynamic ranking (DDR) and the accumulation for dynamic ranking (ADR) as alternatives to the patent TA (or its extended patent TA). Both DDR and ADR consider weights in all dimensions (rather than just the highest and lowest weights as in the patent TA and extended patent TA). Moreover, the consideration of jumping to the furthest document in both DDR and ADR helps reduce the likelihood of checking a sequences of similar irrelevant documents in traditional approaches such as clustering-based geometrical structure retrieval [34.20].

As illustrated in our examples, which are consistent with some empirical results [34.19], ADR usually checks fewer irrelevant patent documents than DDR when finding the top-k relevant patent documents. For instance, to achieve a 100% recall (i. e., to retrieve all relevant patent documents) in a patent search, traditional approaches checked 400 patent documents whereas DDR and ADR checked only 230 patent documents and 80 patent documents, respectively.

34.6 Conclusions

This chapter focuses on information technology (IT) based patent retrieval models. It first highlighted some key differences between patent retrieval and general information retrieval (IR). For instance, patent retrieval is recall-oriented, whereas general IR is precision-oriented. In terms of the search space and time, patent search usually requires longer search time than general IR because (a) patent search queries usually contain more keyword terms than general IR search queries and (b) patent documents are usually longer and more complex than web or news articles searched by general IR. Many existing patent retrieval systems use the boolean model, which was designed for IR. Hence, to facilitate effective patent search, a basic model that was designed for patent retrieval was described. It finds the top-*k* relevant patent documents with no irrelevant documents being returned. The model first extracts representative keyword terms from the user query, then ranks the patent documents containing these keyword terms by assigning weights based on term frequency-inverse document frequency (TF-IDF), and finally returns the top-*k* answers to the user query. Such a model can be enhanced and extended in different aspects:

1. First, the user query can be enhanced by:
   a) Taking a sample and extracting representative terms from the sample
   b) Incorporating relevance feedback about partial knowledge on relevant and irrelevant documents
   c) Estimating the importance of keyword terms.
2. Second, patent documents can be cleaned up by text preprocessing tasks like syntax tagging, word stemming, and stop-word elimination.
3. Third, the ranking of the patent documents can be enhanced by taking into account the international patent classification (IPC) categories.

Hence, patent documents containing the keyword terms belonging to any of these categories can be ranked according to the weights assigned based on the term frequency-inverse patent category frequency (TF-IPCF). The resulting ranked lists of patent documents can be represented as a matrix or a collection of vectors. Moreover, in this chapter, four different top-*k* patent retrieval algorithms were described:

1. Patent threshold algorithm (patent TA) that handles TF-IDF
2. Extended patent TA that handles TF-IPCF
3. Dispersion for dynamic ranking (DDR)
4. Accumulation for dynamic ranking (ADR).

When using these four algorithms in the basic patent retrieval model or its extensions, the top-*k* answers to a user query can be retrieved efficiently with noise elimination, i. e., patent documents that are relevant to keyword terms in the user query can be retrieved efficiently without returning any irrelevant patent documents.

References

34.1 U.S. Patent and Trademark Office: U.S. patent activity calendar years 1790 to the present, https://www.uspto.gov/web/offices/ac/ido/oeip/taf/h_counts.htm

34.2 World Intellectual Property Organisation: International Patent Classification (IPC), http://www.wipo.int/classifications/ipc/en/ (2018)

34.3 A. Cuzzocrea, W. Lee, C.K. Leung: High-recall information retrieval from linked big data. In: *Proc. IEEE COMPSAC*, Vol. 2, ed. by S.I. Ahamed, C.K. Chang, W. Chu, I. Crnkovic, P. Hsiung, G. Huang, J. Yang (IEEE Computer Society, Los Alamitos 2015) pp. 712–717

34.4 S. Adams: A practitioner's view on PaIR. In: *Proc. PaIR 2011*, ed. by M. Lupu, A. Rauber, A. Hanbury (ACM, New York 2011) pp. 37–38

34.5 C.J.V. Rijsbergen: *Information Retrieval*, 2nd edn. (Butterworth-Heinemann, Newton 1979)

34.6 E. Zhang, Y. Zhang: Average precision. In: *Encyclopedia of Database Systems*, ed. by L. Liu, M.T. Özsu (Springer, New York 2009) pp. 192–193

34.7 W.B. Frakes, R. Baeza-Yates (Eds.): *Information Retrieval: Data Structures and Algorithms* (Prentice-Hall, Upper Saddle River 1992)

34.8 M. Kobayashi, K. Takeda: Information retrieval on the web, ACM Comput. Surv. **32**(2), 144–173 (2000)

34.9 A. Lashkari, F. Mahdavi, V. Ghomi: A boolean model in information retrieval for search engines. In: *Proc. ICIME 2009* (IEEE Computer Society, Los Alamitos 2009) pp. 85–389

34.10 W. Lee, C.K. Leung, J.J. Song: Reducing noises for recall-oriented patent retrieval. In: *Proc. BD Cloud 2014*, ed. by J. Chen, L.T. Yang (IEEE Computer Society, Los Alamitos 2014) pp. 579–586

34.11 R. Fagin, A. Lotem, M. Naor: Optimal aggregation algorithms for middleware. In: *Proc. PODS 2001* (ACM, New York 2001) pp. 102–113

34.12 J.J. Song, W. Lee: High recall-low cost model for patent retrieval. In: *Proc. Big DAS 2015*, ed. by C.K. Leung, A. Nasridinov (ACM, New York 2015) pp. 213–216
34.13 C.D. Manning, P. Raghavan, H. Schütze: *Introduction to Information Retrieval* (Cambridge Univ. Press, New York 2008)
34.14 J.J. Rocchio: *Relevance Feedback in Information Retrieval* (Prentice, Englewood Cliffs 1971)
34.15 S. Kullback, R.A. Leibler: On information and sufficiency, Ann. Math. Statistics **22**(1), 79–86 (1951)
34.16 K. Toutanova, D. Klein, C.D. Manning, Y. Singer: Feature-rich part-of-speech tagging with a cyclic dependency network. In: *Proc. NAACL 2003* (Association for Computational Linguistics, Stroudsburg 2003) pp. 173–180
34.17 S. Chakrabarti: *Mining the Web: Discovering Knowledge from Hypertext Data* (Morgan-Kaufmann, San Francisco 2002)
34.18 J.W. Reed, Y. Jiao, T.E. Potok, B.A. Klump, M.T. Elmore, A.R. Hurson: TF-ICF: A new term weighting scheme for clustering dynamic data streams. In: *Proc. ICMLA 2006*, ed. by A. Wani, T. Li, L. Kurgan, J. Ye, Y. Liu (IEEE Computer Society, Los Alamitos 2006) pp. 258–263
34.19 J.J. Song, W. Lee, J. Afshar: Retrieving patents with inverse patent category frequency. In: *Proc. Big Comp 2016* (IEEE, Piscataway 2016) pp. 109–114
34.20 J. Wu, Y. Chen, D. Dai, S. Chen, X. Wang: Clustering-based geometrical structure retrieval of man-made target in SAR images, IEEE Geosci. Remote Sens. Lett. **14**(3), 279–283 (2017)

Carson Leung

Department of Computer Science
University of Manitoba
Winnipeg, Canada
kleung@cs.umanitoba.ca

Carson Leung is a Professor of Computer Science at the University of Manitoba, Canada. He obtained his BSc(Hons), MSc and PhD degrees from the University of British Columbia, Vancouver, Canada. He has contributed more than 180 refereed publications on science and technology, databases, data mining, big data science, social network analysis, as well as data and visual analytics. He won the Best Paper awards in SCA and ASONAM-FAB.

Wookey Lee

Industrial Engineering
Inha University
Incheon, South Korea
trinity@inha.edu

Wookey Lee is a Professor at Inha University, South Korea. He received his BS, MS and PhD degrees from Seoul National University, South Korea, and his MSE from Carnegie Mellon University, USA. His research interests include graphs, anonymization, and patents. He has served as a Chair and PC for CIKM, DASFAA, VLDB, and BigComp. He won the Best Paper awards in IEEE TCSC, KORMS and KIISE.

Justin JongSu Song

Industrial Engineering
Inha University
Incheon, South Korea
jaegal83@inha.edu

Justin JongSu Song is currently a PhD candidate at Inha University, South Korea. He received his BSc and MSc in Industrial Engineering with high honours from Inha University. His research interests include graph theory, social networks, information retrieval, and patent analysis. He won the Best Paper awards in KIPA, IEEE TCSC, KORMS, and KIISE.

35. The Role of the Patent Attorney in the Filing Process

Rainer Frietsch, Peter Neuhäusler

The role of the legal representative in patent filing processes is, so far, under-explored in patent statistics. This chapter addresses the question of the role and the impact of the patent attorney in the filing process. One of the core assumptions is that more experienced attorneys have more in-depth knowledge of the intricacies of the patent system and, thus, are more likely to pursue more elaborate and successful filing strategies.

The results show a high concentration of attorneys and filing action in absolute as well as in relative terms in some countries, namely Germany and the UK, and numbers worth mentioning also in other larger applicant countries like France, Italy, Sweden, or the Netherlands. Explanations for this biased distribution in Europe are language advantages in the case of the UK (and also Ireland) and geographical proximity to the European Patent Office (EPO), as well as economies of scale in the case of Germany.

The experience of the representative has a considerable impact on the outcome. Multivariate analyses suggest that the (financial) resource endowment is a decisive factor in the hiring of patent attorneys. It was shown that the patents of more experienced representatives were significantly more often withdrawn (but neither refused nor granted with a higher probability), and they were less often opposed than the ones by less experienced attorneys.

35.1 Starting Points

In the past 20 years, the patent system underwent a considerable change in what applicants/assignees expect from it and how they use it. Though still being important, the pure protection mechanism lost relevance compared to the more strategic aspects like blocking competitors or signaling to the capital market [35.1]. Though strategies are people's reactions to the expectations on future developments, the role of a certain group of people, namely patent attorneys, has mainly been disregarded in the literature and is—in consequence—rather under-explored so far. People are involved in the production, drafting, and especially decision making on patents. Consequently, there is a human factor in the patenting process that plays a crucial role. The applicant's representative and the patent examiner are usually in direct contact and communicate on the different stages of filing a patent. The examiner provides the applicant (or his representative) with information on the application and might make suggestions or even requests to the representative, who, on the other hand, is able to react to this in one way or another. In other words, patenting is a process where negotiations play a major role.

Patents are a vested right for a limited time, which applicants are granted. Patents are essential to secure the investment in research and development (R&D) and in the technological progress, especially in innovation-oriented countries. Applicants are granted this right in exchange for the disclosure of their technology/process. Furthermore, patents have to fulfill several formal criteria, among them novelty and inventive steps, which are both examined by professionals in the particular field. The formulation of patent applications is a sophisticated and challenging task, mostly fulfilled

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_35

by professionals—patent attorneys—who then take responsibility of (in most cases) massive investments in R&D to be protected. A patent application is not just a simple matter of filling in a form. The way from filing to the granted patent is a long one and is full of obstacles and traps. All this makes the patent attorney—or more, generally speaking, the representative—a very important person within this process. High qualifications, broad experience, and a number of other characteristics are mandatory, or at least helpful, for a successful patent application.

For the representation at the EPO there are clear rules and regulations on who might undertake this legal act on behalf of the patent applicant. Article 134 of the European Patent Convention (EPC) deals with the question of who could act as a representative of the patent owner [35.2]. This can only be done by a natural person and a national of one of the 38 EPC member states, who has a place of business within the member states and has passed the European qualifying examination. These are so-called European patent attorneys. In addition, any legal practitioner entitled to represent patent owners in patent matters in any of the EPC member states might also be able to act as a representative at the EPO. So, any national patent attorney could also act on behalf of the patent applicant.

In addition to legal practitioners (lawyers/attorneys), a patent applicant who has their place of business—this means either headquarters or a subsidiary organized as its own legal entity—within the EPC member states might also file a patent without appointing a legal practitioner. Even non-residential entities who have their place of business outside the territory of the EPC might file an application on their own account, but then need to appoint a legal practitioner in the further course of the process.

In the guide for applicants [35.3] the EPO strongly recommends appointing a professional legal representative, especially to keep the probability of a successful application high. Different to the US, where individual applicants who act pro-se, i. e., on their own, get support and even consultancy from the examiner; such a special rule does not exist at the EPO. So, if an individual fails to fulfil his/her requirements of the process, the filing procedure might end, and the patent be rejected without any benefit of doubt. In the US, the applicant might even get some support and consultancy on how to reformulate some claims or change the scope of the patent to keep the process running. However, also in Europe a communication between the examiner and the applicant (or his representative) occurs. The applicant might derive some useful information on how to refine his application, but there is no formal support for unexperienced applicants.

The majority of patent applications are filed by a professional legal representative, who can either be internal—so, directly employed by the patent applicant—or external. Especially small and medium-sized enterprises (SMEs)or at least small applicants (applicants with only few patent applications) might choose external representatives, whereas larger enterprises and larger applicants might be able to sport their own patent attorney or even their own legal department with more than one attorney. However, in reality, large multinational enterprises with their own legal departments employ both internal and external representatives, depending on the workload and the sensitivity of the particular application. It is often the case that a number of different legal practitioners are involved in a single patent filing procedure. Even within the patent system there is plenty of room for specialization. One person might be an expert in drafting the text, another one even in writing the claims, a third one in corresponding with the office, and another one being the best to act at the court of appeals in the case of opposition or in law suits when already granted patents are under attack. For licensing and contracting yet another legal representative might enter the scene. In fact, also in most smaller law firms and certainly in larger ones, a patent application is usually teamwork, where prior art searches, correspondence, process management, fee payment, etc., are done by a team of people. Nevertheless, in the patent application, only one person is mentioned as having to fulfill the criteria of Article 134 mentioned above. This particular person named on the patent as a representative is the scope of this analysis.

Once the decision to hire a lawyer was made, one might employ an experienced or a less experienced attorney. Larger companies might even employ internal legal staff and then have the choice between an internal and an external representative. Of course, applicants who frequently file patents might prefer to have permanent staff that is familiar with the technologies, the strategies, and the habits of the company. Less frequently or irregularly filing assignees might not have these options, or their lawyers might not only do patents but also fulfill other legal duties. At least, they might not have the same opportunity to gain experience in patenting. The same holds for external representatives, who might be specialized in patenting, filing hundreds of patents per year. Others might also be representatives for other legal issues, therefore filing less patents per year and also having less experience with patents. It is necessary to register with the U.S. Patent and Trademark Office (USPTO) to be allowed to act in front of the USPTO. The same holds for the EPO, where it takes about 3 years of training to become a European patent attorney. In sum, the individual skills and experience of the lawyer and the law firm, as well as the framework conditions might have an impact on the outcome of the filing.

The costs of filing patents is in most cases defined by the procedural costs of the endeavor, where processing fees at the patent offices are calculable and straightforward and are, therefore, often used to assess the overall costs of a patent application. However, they are only a part of these overall costs. The costs for the services of the patent attorney, for translation, for prior art searches, etc., are also considerable and make patents a costly thing [35.4, 5]. It becomes even more costly—though hard to predict—when law suits or infringements are involved. Moreover, this holds for both parties of law suits or infringements. This is why the resource endowment is often discussed as a crucial factor for pursuing one's own patent rights. Further, the resource endowment might also extend to the legal representative, where a *better* one might be more costly than the *average* one, and this might have an impact on the outcome of the filing process.

The role of the legal representative in this procedure is a so far an under-explored research question in patent statistics. This chapter addresses the question of the role and the impact of the patent attorney on the filing process. One of the core assumptions is that more experienced attorneys have more in-depth knowledge of the intricacies of the patent system and, thus, are more likely to pursue more elaborate filing strategies.

This chapter has the following structure. A brief literature review on the scarce empirical and theoretical evidence describes the framework conditions under which the role and the impact of the attorney on the filing outcome are analyzed. Next, the dataset to describe the structures is introduced, and a descriptive statistical section is followed by the discussion of the multivariate empirical results. A final section summarizes and concludes.

35.2 Literature Review and Regulations

The importance and the role of the patent attorney has been mentioned in several publications [35.6–9], but the empirical evidence that examines this impact is scarce, and quantitative empirical evidence hardly exists. One of the few papers that contains empirical evidence was presented by *Somaya* et al. [35.10], who stress the skills and competencies of the attorneys in the patent process or in drafting patents.

Macdonald and *Lefang* [35.11] have a rather skeptical perspective on the role of the patent attorney in producing innovations. However, already two decades ago they saw an increasing role of the representative for the invention activity. The reasons were an increasing importance of the patent system in general, as well as tacit knowledge within the system that was supposed to gain relevance. They conclude that the system, thereby, deviated from its original and social task to help securing innovations.

Li et al. [35.12] stress the potential role of patent attorneys as intermediaries or match-makers for technology transfer in particular in developing countries. They argue that inventors who have the technological expertise often miss competence in commercialization and in identifying partners for collaboration. On the other hand, companies in need of technologies often lack the knowledge of what is technologically possible and what already exists. In the discussion of *Li* et al. [35.12], the patent attorneys are seen to close this gap and act as match-makers in these kinds of situations. However, what they neglect in their line of argumentation is that not seldom companies are suffering from the not invented here (NIH) syndrome, which means that they are unable to absorb or even accept technologies from outside the company [35.13–15]. In addition, not seldom the technological solutions suggested by individual inventors or public research is of no direct use but requires additional investments in development activities to get them ready for the market, to adapt them to the requirements of the interested company, and to enable their upscaling for mass production.

One of the few papers that examines the impact of the attorney on the outcome of filings was published by *Gaudry* [35.16], who analyzed the difference between attorney representation and pro-se applicants at the USPTO. She found considerable differences not only in the success rate of the application as such, but also in terms of the number and broadness of claims. Pro-se applicants do not fully exploit the possibilities of the system, and it appears that several of them abandon the process unintentionally due to a lack of experience with the legal issues and the procedures.

A paper published by *Koller* and *Ebersberger* [35.17] uses the experience of patent attorneys, measured in years since their examination at the EPO or as a cumulated number of previously represented patents, to analyze their impact on the filing outcome. They measure the outcome as the duration from application to grant, as well as the number of forward citations received. They, indeed, find a positive relation between experience and time to grant and also the number of forward citations, which is an indication of patent quality in their eyes. However, what they are not able to take into account in their models is the intentional or strategic use of the legal status [35.1, 18]. In some cases, a withdrawal or even a refusal is sufficient to achieve one's goals. In other words, not all patents that are not granted are of no or low value, whereas also not all patents that are granted are of any value at all [35.19]. Moreover, it could very well be that experience enables attorneys to act strategically in this way.

A few papers also address the impact of the judges and the courts in case law systems. For example, *Moore* [35.20] starts from the widespread perception of the existence of jury bias in patent trials in favor of individuals and in disfavor of corporations. She, indeed, finds empirical evidence for the existence of such a bias as individuals win 78% (companies win 22%, respectively) of cases in a jury trial, while they only win 47% in trials decided by a judge [35.20, p. 82]:

> Asymmetries in resources, stakes, and information, differences in risk preferences, concern over adjudicator bias, behavioral science, and repeat-player economies all suggest a difference in the pool of cases litigated by individuals as opposed to the pool litigated by corporations.

This leads her to conclude that there is a selection bias at play at the same time when cases are tried. Corporations might only try the solid ones, while individuals might not be able to assess this in advance. She even finds empirical evidence for the statement that individuals do not file fewer law suits because they are more selective and more risk averse, but because their patents are—on average—not as strong as the ones held by larger parties. She explains this by the affection of the individual for her own invention. In addition, she refers to anecdotal evidence and argues that individuals who are not afraid of trouble might be more willing to file weak law suits. On the other hand, larger parties might use their resources to file litigations against smaller, less resource-equipped patentees and tear them down or reach a settlement simply by the power of their resources.

Reitzig and *Wagner* [35.21] derive from a number of interviews that external patent attorneys are—due to possible economies of scale—more efficient in filing patents. However, in the case of litigation, internal attorneys are more efficient and effective according to their findings.

Apart from patents, the fact that the attorney—or to be more precise: her/his ability and experience—has an impact on the outcome of legal procedures was shown in several papers. For example, *Abrams* and *Yoon* [35.22] use felony cases randomly assigned to attorneys to show that experience matters—and that this experience is not equally distributed among attorneys [35.22, p. 1146]:

> In civil cases, higher-ability plaintiff attorneys are more likely to win, and garner larger damage awards for their clients; higher-ability defense attorneys are more likely to avoid a finding of liability for their clients, or at least minimize the size of the damage award

Also *Szmer* et al. [35.23] take the capabilities and experience of the lawyer into account when they analyze Supreme Court decisions in Canada. They find that litigation experience and team size have an impact on the performance.

35.3 Basic Research Questions

The analytical potential of our dataset regarding patent attorneys has many dimensions (the data will be presented below), so in this chapter, we restrict ourselves to some basic questions as the existing literature does not provide satisfying answers even to these basic questions. Does the experience of the patent attorney differ by applicant type (small and midsize enterprise, large enterprise, university, public research, or individual inventor)? Does the experience of the patent attorney have a significant impact on the legal status of a patent? Additional questions, could be, whether more experienced attorneys achieve more valuable patents or are hired to process more valuable patents (measured by forward citations, according to International Patent Classification (IPC) classes, or claims)? Are patents drafted by experienced attorneys opposed less often? Finally, the question of all questions in this context is—and where we will be able to also only deliver a few indications—whether more experienced attorneys are more capable of strategic patenting than less experienced colleagues?

35.3.1 Data

This chapter uses a dataset of more than 1.8 million patent applications filed at the EPO between 1990 and 2010—we had data up to 2012, but the Patent Cooperation Treaty (PCT) filings of the last 2 years had not yet entered the regional phase at the EPO by then. However, for country comparisons, we use only data from the latest available year, while we use data of the whole period for the multivariate analyses. The data was extracted from PATSTAT, also known as the EPO Worldwide Patent Statistical Database, which contains mainly bibliographic information extracted from the INPADOC (International Patent Documentation) database that is used by the patent examiners of the EPO and which covers patent filings at more than 80 offices worldwide [35.24, 25]. The database is run on an Oracle SQL Server, which offers high potential to calculate and introduce additional indicators. Next to standard data, our database also contains the information on the size [35.26] and type of the assignee, so if it is an SME, a large enterprise, a university, a public research organization, or an independent inventor. Assignee types have also been identified by the K.U. Leuven [35.27–29], which can be added to PATSTAT as well. However, this data was not used here, as the most relevant information on the size of the applicant is not available. We, therefore, used our own definition of applicant types, which, however, is only available for nine countries.

Furthermore, the PATSTAT database was enriched and matched with additional information, like the WIPO (World Intellectual Property Organization) classification on technological fields [35.30]. In addition, for the sake of this paper, the PRS-data was included, which contains the legal status of the patent filings. This data source offers information for several patent offices if a patent is granted, refused, withdrawn, pending, or if the granted patents are maintained. Another product of the EPO that is essential in the context of this paper is the *ESPACE Bulletin*, where the name and address of the legal representatives—covered in addition to several other information already available in PATSTAT—was extracted and added to our database. After the extraction, the data was processed, and each law firm received an individual ID as did each patent lawyer. As many patents only contain information on the law firm, and the individual lawyer/representative is not always mentioned, the coverage of this latter variable is much lower than that of the law firm. Therefore, we restrict our analyses on law firms in this chapter.

In some cases—especially large enterprises—the legal department, an IPR (intellectual property rights) subsidiary and/or individual persons were named. They were then directly coded as internal representatives. In cases where no representative was mentioned at all, they were coded separately, but also treated as if they were internal representatives.

The experience of the law firm or the individual lawyer in a particular year was calculated as the cumulated number of EPO patent filings up to that particular year, starting in 1978, when the EPO started its business. It was calculated this way as this is system immanent information. One could argue that the law firms and lawyers might file patents outside the EPO and, thereby, also gain experience, for example, in drafting a patent or negotiating with the office. However, as the analysis is restricted to EPO patent applications, the experience within the system seems to be a legitimate and appropriate perspective.

Experience increases over time. It does so as a matter of fact and as a matter of the construction of this indicator, which is defined as the cumulated number of filings over time. It is reasonable to assume decreasing marginal effects of individual filings on the additional experience. In other words, the first and the second patent filing might add much more to the overall experience of a law firm or an individual lawyer, while the 101st or the 1001st filing will add much less. To take this effect of decreasing marginal effects into account, the logarithm of the cumulated number of filings is used instead of the absolute number. However, there is still the problem of strictly increasing experience over time due to the construction of the variable, which might have an effect on the estimations. Moreover, even if the time is controlled by using the information on the filing year, there is a specific impact due to the fact that the total number of filings and, therefore, the level of additional experience are not constant over the years. While the number of filings was much lower in the 1980s, they steeply increased in the second half of the 1990s and—more or less—stabilized on a higher level in the 2000s [35.1, 31, 32]. To cope with this, the individual experience of a law firm or a lawyer, respectively, could be benchmarked against the average experience level across all actors in the patent system in that particular year. This indicator could be characterized as a relative experience level in a particular year, but independent of time effects. However, for the sake of simplicity, and as the overall results allow the same conclusions, we restrict our empirical analyses to the former variable and use the logarithm of the cumulated number of patent filings as the indication for experience here.

35.4 Descriptive Results

All the representatives are located in one of the EPC member states. However, we restrict our focus to EU-28 member countries here. The total and relative numbers (in relation to the number of inhabitants to control for size) vary considerably between countries. Focusing on the year 2010, where we have complete data including PCT filings that already entered the regional phase at the EPO, the numbers of representatives in the member countries vary considerably between almost 2400 in the case of Germany and none at all in the case of Malta (Fig. 35.1a). We count about 1300 representatives in the year 2010 in the UK and about 570 in France, which ranks third. In total, we were able to identify 6286 different representatives in the year 2010.

The ranking changes completely when size effects of the countries are taken into account by relating the number of representatives to the countries' population (Fig. 35.1b). In this perspective, Luxembourg ranks first with almost 40, followed by Germany with 29. The Scandinavian countries Sweden and Denmark are slightly ahead of the UK, all of them having more than 20 attorneys per 1 000 000 inhabitants. Countries like Austria, France, Ireland, or Italy reach numbers between 5 and 10.

In this section, we start with the analysis of the average number of patents previously filed by the representatives—calculated as the accumulated number of patent filings for each representative (Fig. 35.2). One factor influencing the experience is the duration of practice. Eastern European countries have a disadvantage here due to the later accession not only to the EU, but also to the EPCT that underlies the EPO. While the EU average of years of practice per representative is 12, that of eastern European countries is much lower, but also Spain or Luxembourg show below average values. There are a couple of countries that are close to this average or even above, namely with the UK, Ireland, and Austria ahead and Germany, Italy, and the Netherlands slightly below.

The shares of most of the countries in the period between 2000 and 2010 are rather stable for most of the countries with slightly increasing trends in Germany

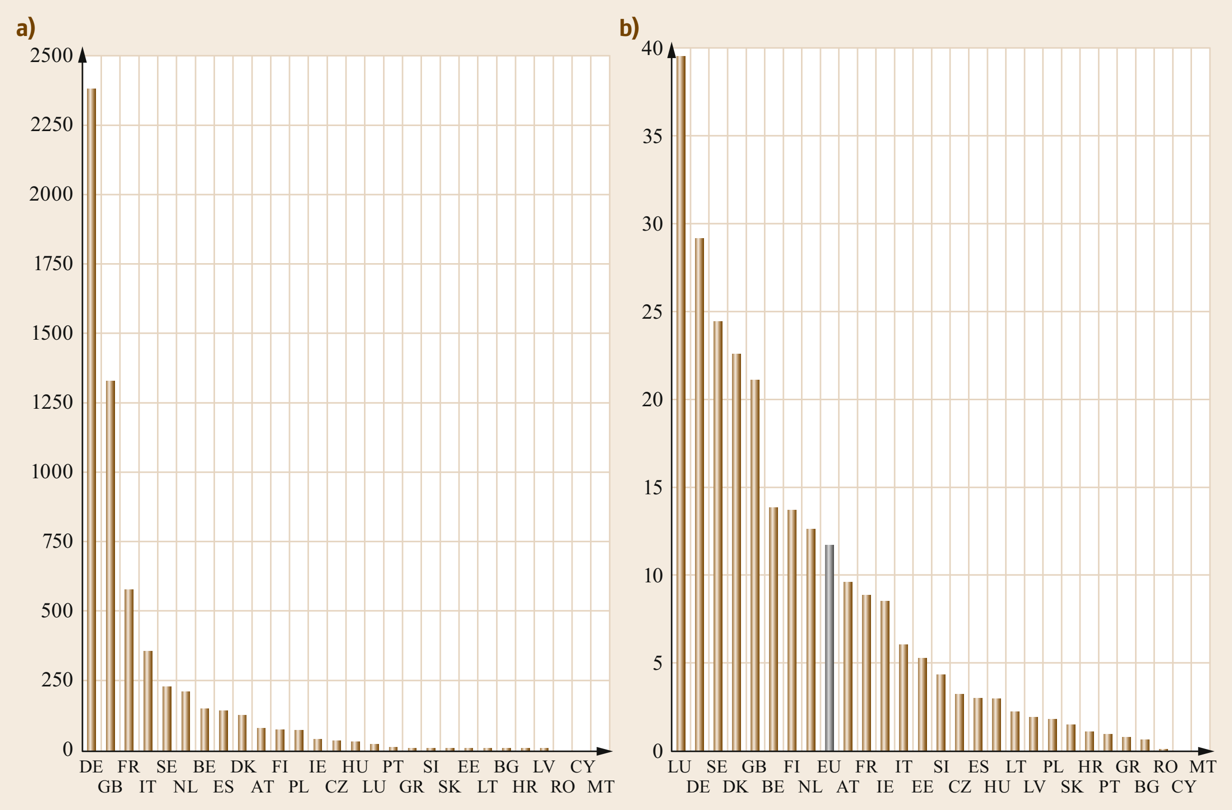


Fig. 35.1a,b Total number of representatives **(a)** and in relation to 1 million inhabitants **(b)** by country of residence of the representative, 2010. Source: EPO-PATSTAT; Fraunhofer ISI calculations

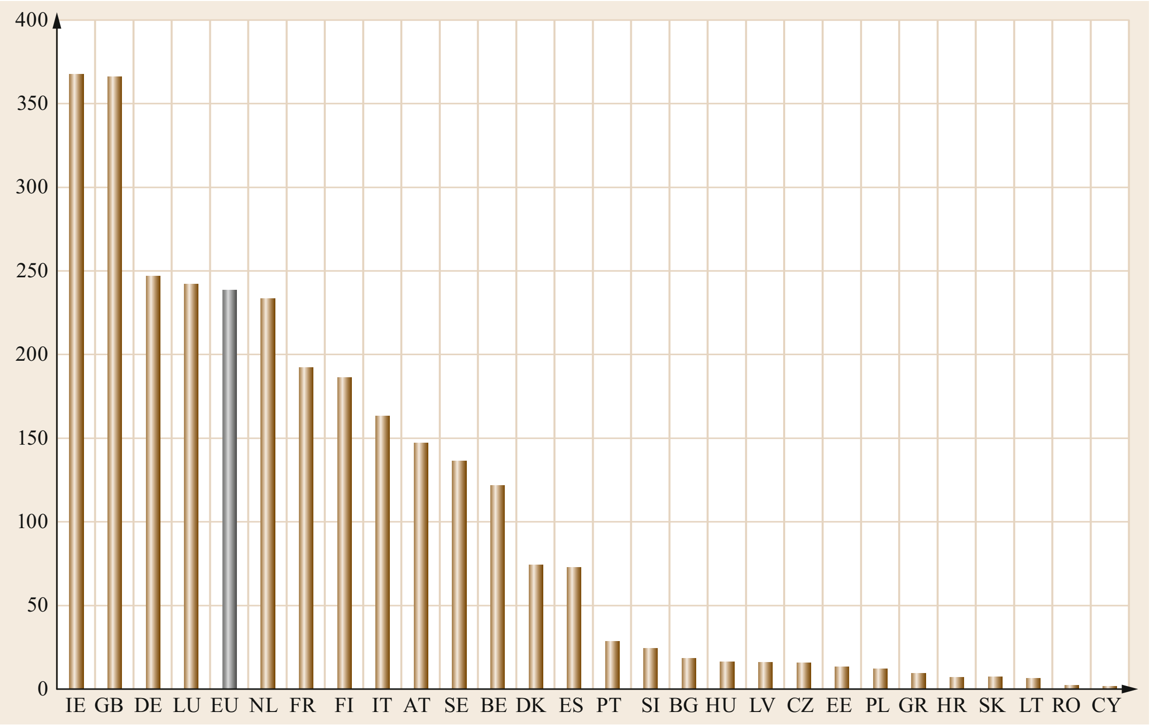


Fig. 35.2 Average experience (cumulated number of filings) of representatives by country, 2010. Source: EPO-PATSTAT; Fraunhofer ISI calculations

and France, but clearly a decreasing trend from 32.4 to 25% in the UK. However, on average, a representative in the UK and in Ireland previously represented almost 370 patents. The numbers for Germany, Luxembourg, and the Netherlands, as well as the EU average accumulate to almost 250, whereas in most other western European countries like France, Finland, and Italy the experience sums up to below 150 patents.

The explanation for these considerable differences is that the high share of the patents processed by UK, Irish, and also by German representatives are filings from non-European applicants, which are subsequent filings to priorities outside Europe. In many cases, these documents are prepared by legal practitioners outside Europe. As these non-European practitioners are not allowed to file patents at the EPO, a corresponding attorney from Europe is necessary to act in between. Essentially, these representatives have much less work, as they often only act as a corresponding representative, so that they can process (in our case simply occur on) many more patents. This is much less work, but also results in less experience in drafting patent applications than if they had to fully prepare them from scratch. Obviously for practical reasons, UK and German representatives are chosen rather frequently by non-European applicants in absolute terms. These practical reasons mainly stem from language issues on the one hand and sheer absolute numbers (availability) on the other hand, but in the end also from experience (economies of scale) as a kind of self-fulfilling process. In relative terms, however, there is a big difference between the UK and the rest of the European countries. As an indication of this, the relation between patents processed by national representatives and the number of EPO filings by applicants from each country is calculated. The most outstanding data can be found in the case of the UK. Representatives from the UK processed about seven times more patents in 2010 than there were filings from UK applicants (Fig. 35.3). The language advantage seems to be the decisive factor here and makes UK law firms the main entry point for US-American applicants—and of course there are also several *European headquarters* of non-European multinational companies in the UK, which is itself partly also a consequence of the language. The effect might be amplified by the fact that the UK industry is not very patent-intensive and, therefore, files rather low numbers of patents per employee [35.32] so that applicants (and filings) from the UK themselves are rather scarce. Inventors from the UK file about 250 patents per one million employments, whereas, for example, inventors from Germany file 770 patents per one million employments [35.33].

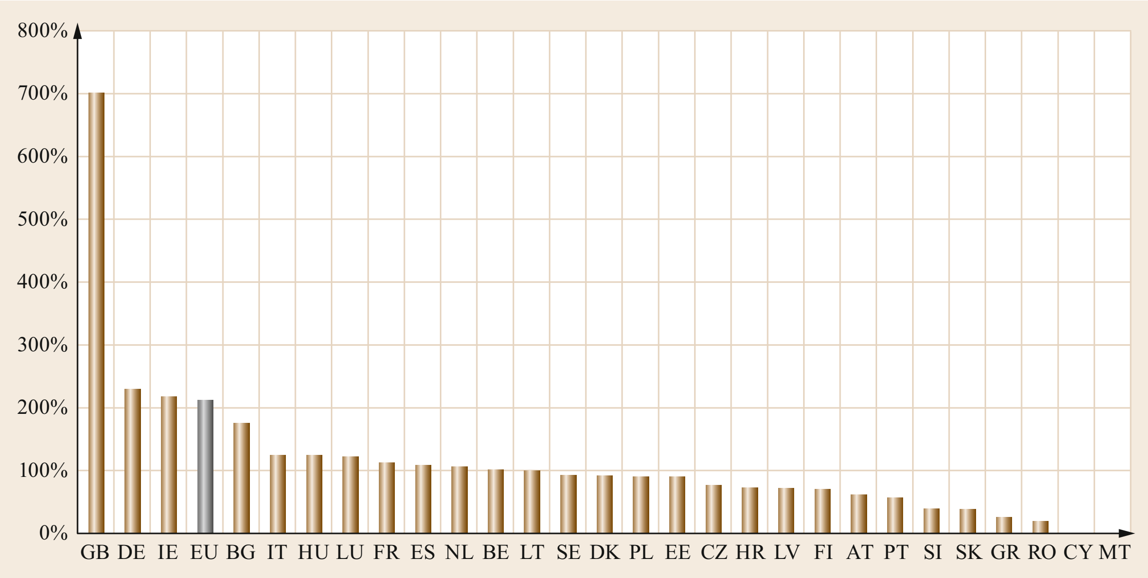


Fig. 35.3 Number of patents processed by representatives in a country as a share of applications from that particular country. The numbers for Bulgaria (and several other smaller countries) should not be over-interpreted. These are based on only 14 patents processed by Bulgarian representatives in 2010). Source: EPO-PATSTAT; Fraunhofer ISI calculations

In Germany and Ireland the relation is more than 2 : 1, with Ireland benefiting from the language advantage as well (including the *European headquarter* effect). Germany is the largest applicant country from Europe at the EPO and files rather high numbers of patents in relation to its work force (patent intensity). It is, therefore, even more remarkable that two times as many patents are processed by German representatives than there are filings by domestic applicants. The explanation for this is—again, in addition to the *European headquarter* effect—simply the large number of law firms, resulting from the proximity to the EPO, but also from the high domestic demand for patent practitioners.

Figure 35.4 provides additional evidence that supports these explanations. Here, a change of the level of analysis from the number of representatives to the number of patents is made. The figure shows the share of patents of three applicant country groups—domestic, other-EU, and countries outside Europe—by the country of residence of the representative.

It can be seen that representatives in the UK, Ireland, Germany but also Bulgaria and Luxembourg process high shares of applications by non-European applicants. Slovenia's and Slovakia's representatives only process patents for domestic applicants. In countries like Finland, Estonia, the Czech Republic, or also Austria, the representatives process high shares of applications by domestic applicants. High shares of filings from applicants located in other European countries can hardly be found. Greece, Hungary, and the Czech Republic, as well as Latvia are exemptions, but with low absolute numbers; 12.4% of the patents processed by Belgian representatives are owned by applicants from other European countries, which is a rather high share compared to most of the other countries. This might be due to the language advantages of Flemish representatives, also processing filings from the Netherlands and Wallonian representatives processing filings from France.

This general finding is further supported by the applicant structure in the countries. German and UK representatives file in about two-thirds of the cases patents for large enterprises. Also, in the Netherlands, Finland, and Sweden rather high shares of larger enterprises' patent filings can be found, which is the result of one or a few very active enterprises in these countries. In France, the shares are similarly high with about 55%. In eastern and southern European countries, on the other hand, there are almost no large enterprises to be found, but higher shares of SME, and often individuals and public research organizations account for shares of around one fourth of all filings. In Greece, for example, half of all filings processed by domestic representatives stem from individuals.

In Belgium and the Netherlands about 40% of patents are represented by internal attorneys—as a reminder: both countries have high shares of large enterprises of more than 50%. Also in Finland or Sweden, the share of internal representatives is close to 30%. In the UK and Germany, this share is about 6.5 and 6.3%, respectively. At least in the case of Germany, this might be somehow surprising, as the share of patent applica-

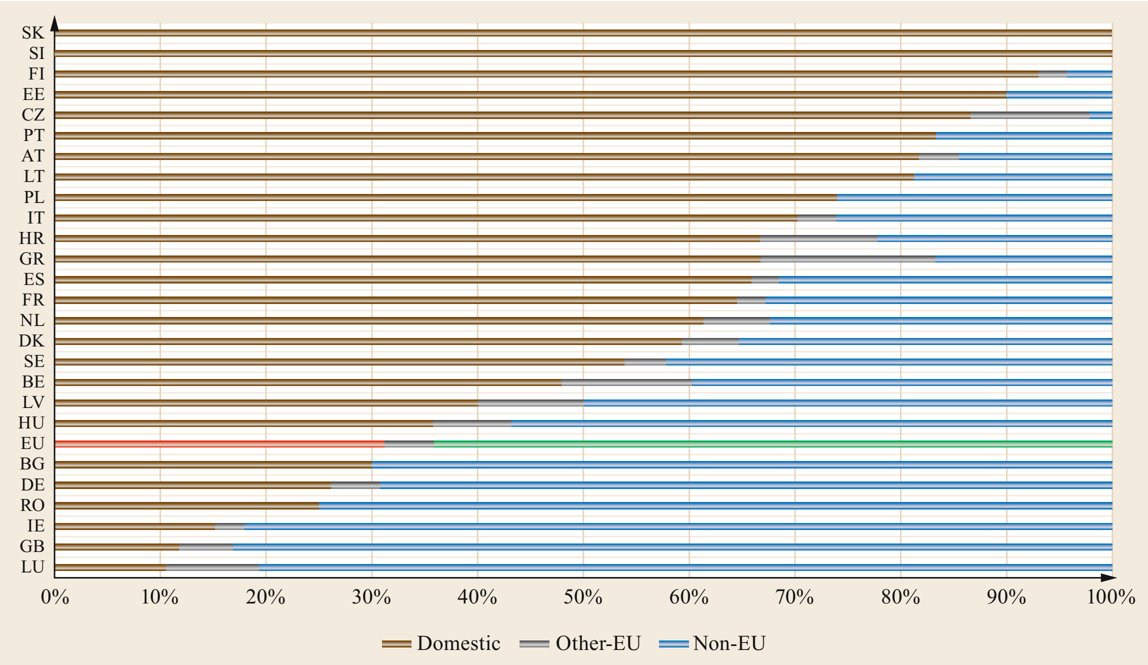


Fig. 35.4 Shares of applicant country groups by country of representative, 2010. Due to low absolute numbers, data for Malta and Cyprus cannot be analyzed. Source: EPO-PATSTAT; Fraunhofer ISI calculations

tions filed by large applicants is rather high [35.26]. As large applicants have a higher probability of employing internal representatives, the expectation for this share might have been higher. One explanation again is that about two-thirds of the patents processed by German representatives are filed by non-European applicants. Another explanation is that there might be a higher division of labor between internal and external representatives, simply due to the fact that German companies have a high absolute output of patent filings, which cannot all fully be processed by internal representatives.

In general, our data suggests that large enterprises employ far more internal representatives than any of the other groups. It is interesting to note, however, that the shares of internal representatives are higher for applications to the EPO by non-EU applicants than by applicants from EU-28 countries. The explanation might be that a number of multinational companies have subsidiaries in Europe, very often with intellectual property (IP) departments or even simply IP management subsidiaries, which then fulfil the criteria of Article 134 of being a resident of the EPC states and/or having its place of business in Europe.

The average number of technological fields per representative in each of the countries of residence of the representative is shown in Fig. 35.5. Representatives from Ireland and the UK cover—on average—more than 25 or more than 20 fields, respectively. This might be seen as another indication of rather high shares of subsequent patents with priorities outside Europe, especially in the US, where representatives from these countries mainly act as corresponding representatives. To act as a corresponding representative it is not absolutely essential to be an expert in the particular technological field. At first sight, however, the result for Germany is in contradiction to this line of argumentation, as the average number of fields is only 13. We found high shares of filings for German representatives originating in non-European countries. Based on these additional insights we might be able to further qualify the above-mentioned result. For Ireland and the UK, the language advantage seems to be the decisive factor for the high shares of non-European filings. For Germany, the language advantage does not really hold, but the proximity to the EPO and the economies of scale, due to a large number of national filings might be attracting factors. This then also results in a higher specialization in terms of the average number of technological fields per representative. In general, Fig. 35.5 shows that representatives in the northern and western innovation-oriented countries seem to be able to cover a larger number of technological fields, whereas—maybe also as a matter of low absolute numbers—representatives in southern and eastern European countries concentrate their technological expertise more.

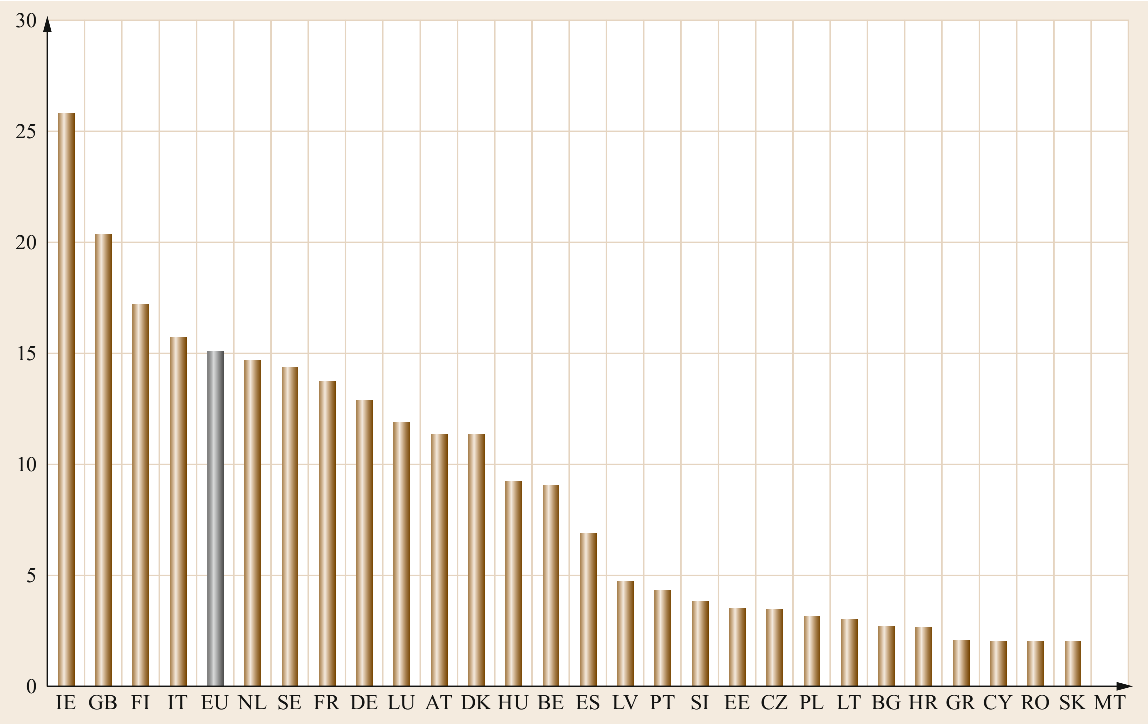


Fig. 35.5 Average number of technological fields per representative by country of representative, 2010. Due to low absolute numbers, data for Malta and Cyprus cannot be analyzed. Source: EPO-PATSTAT; Fraunhofer ISI calculations

35.5 Multivariate Results

The descriptive statistics provide an overview of the structures and allow some conclusions on how these structures come into being. However, the question of the role of the patent attorneys (or more precisely of the representatives) in the filing process is not sufficiently addressed by these statistics. In the following, we provide some multivariate analyses based on the data set we have compiled to address these questions.

The first question we analyze addresses whether large enterprises are able to hire more experienced representatives. We will control for some other features like internal versus external representative, time, technology field and country (of applicant).

With classical human capital theory [35.34–40] we assume that individuals can increase their human capital endowment by formal qualification and training, but in addition also simply by their doing, whereby they gain experience. The (labor) market rewards higher human capital endowment by higher (average long-term) income. Usually, when the economic returns on human capital investment are to be estimated in empirical research [35.41, 42], the discussion focuses on formal qualification and takes experience as a kind of control into account. Here, we argue the other way around, as all patent attorneys/representatives should have similar formal qualifications; they differ in their experience, which then is rewarded by the market. We do not, however, have information on the rewards in economic terms, but just assume that this is so. What we can observe in our data, though, is the difference in the experience and the structures of the patent applicants employing these more or less experienced representatives. In addition, and this will be the second part of this multivariate analysis, we can also examine the performance of the representatives, according to their experience and given some structural information on the applicant and the patent.

In Table 35.1 the logarithm of the experience is regressed on the applicant type, the year, a dummy that covers internal versus external representatives, as well as field and applicant countries as control variables. We see that all coefficients are statistically significant, and the model explains about 23.5% of the variance in the data, which is a satisfying level. In the case of applicant type, large enterprises are the benchmark. The representatives hired by individual inventors have the

Table 35.1 Ordinary least squares (OLS) regression of log of experience on applicant type, year, and internal representatives

dV: Experience of the representative (log)	Coefficient	SD	*t*	*P* > *t*
SME	−0.419	0.004	−100.91	0
Individual inventor	−0.571	0.008	−75.88	0
Public research organization	−0.171	0.010	−17.84	0
University	−0.258	0.010	−26.97	0
Internal representative Yes/No	−1.142	0.005	−215.49	0
Year			Yes	
Country controls			Yes	
Field controls			Yes	
_cons	5486.282	182.501	30.06	0

Number of obs. = 1 344 366; $F(28, 1344337) = 14\,767.80$; Prob. > $F = 0.0000$; $R^2 = 0.2352$; Adj. $R^2 = 0.2352$
Source: EPO–PATSTAT, ESPACE Bulletin; Fraunhofer ISI calculations

Table 35.2 Logistics regression of the opposition of a patent on log of experience of the representative and a number of additional controls

dV: Patent opposition (yes/no)	Coefficient	SD	*z*	*P* > *z*
Experience (logarithm)	−0.041	0.004	−10.3	0
PCT application No/Yes	0.052	0.016	3.24	0.001
Forward citation count	0.025	0.001	38.96	0
SME	0.028	0.019	1.46	0.145
Individual inventor	−0.236	0.037	6.44	0
University	−0.558	0.055	−10.23	0
Public research organization	−0.451	0.061	−7.4	0
Year			Yes	
Country controls			Yes	
Field controls			Yes	
_cons	29.290	1.939	15.11	0

Number of obs. = 402 773; LR $\chi^2(29) = 8948.06$; Log likelihood = −78 390.696; Prob. > $\chi^2 = 0.0000$; pseudo $R^2 = 0.0540$
Source: EPO–PATSTAT, ESPACE Bulletin; Fraunhofer ISI calculations

least experience, followed by representatives of SMEs. Universities, and even more so public research organizations (PROs), employ only slightly less experienced representatives than large enterprises do. On average, internal representatives have slightly less experience, but this might be due to the fact that large numbers of PCT filings entering the European patent system are just to be represented. This is usually not done by internal attorneys—as by definition these companies do not have a subsidiary or branch in Europe. Moreover, even large companies with subsidiaries or branches in Europe might not have legal departments here.

A multinomial logit model (not displayed) restricted to the already processed patents—so essentially excluding still pending patent processes—with the legal status as dependent variable and the experience of the representative as relevant explanatory variable as well as a number of controls shows no significant difference between granted and refused patents. Only the withdrawal rate increases significantly with increasing experience. This could suggest that more experienced representatives act more strategically, given that withdrawals are seen as a—surely imperfect—indication of strategic behavior. It could also well be just an effect of the PCT filings, which add to the experience of some representatives, without requiring the full workload of drafting a patent. This, however, could then also be seen as a kind of strategic behavior. Not all PCTs are strategic patents, but most strategic patents—defined as patents with postponed decision, rather with option values than direct economic values, and in some cases also with more passive or blocking motives—are PCTs. Filings from the US, most of them entering the EPO via the PCT system, have a much lower probability of being granted and, therefore, a higher withdrawal rate than filings from most other countries [35.18].

The impact of the experience of the representative on the probability of being opposed, given that the patent is granted, is displayed in Table 35.2. Indeed, according to these estimations, the experience of the representative mentioned on the patent signifi-

Table 35.3 Poisson regression of the number of claims on the log of the experience of the representative and a number of controls

dV: Number of claims	Coefficient	SD	*z*	*P* > *z*
Experience (logarithm)	0.012	0.000	77.83	0
PCT application No/Yes	0.620	0.001	1028.56	0
Forward citation count	0.010	0.000	600.16	0
SME	− 0.031	0.001	− 40.16	0
Individual inventor	− 0.102	0.001	− 71.17	0
University	0.021	0.002	11.71	0
Public research organization	− 0.096	0.002	− 49.65	0
Year	Yes			
Country controls	Yes			
Field controls	Yes			
_cons	20.149	0.050	406.55	0

Number of obs. = 1 346 266; LR $\chi^2(29) = 2\,046\,238.01$; Log likelihood = $-8\,451\,727.7$; Prob. > $\chi^2 = 0.0000$; pseudo $R^2 = 0.1080$
Source: EPO–PATSTAT, ESPACE Bulletin; Fraunhofer ISI calculations

cantly reduces the probability that a patent is opposed. Granted patents that reached the EPO via the PCT route have a higher probability of being opposed, as well as such patents that have a higher number of forward citations. While patents filed by public research organizations are less often opposed than those of large enterprises, patents by SMEs have a higher probability of being opposed, given the model specification we chose here and which reaches a satisfying pseudo R^2 of about 5.4%.

The impact of experience on strategic aspects of patenting can also be found in Table 35.3, where we report the results of a Poisson regression of the number of claims. More experienced representatives file significantly more claims: PCT filings—on average—also have a higher number of claims, and university filings have slightly more claims than those of the applications of large enterprises, while SMEs, public research organizations, and also individual inventor's patents are filed with slightly fewer claims.

35.6 Summarizing Discussion

This chapter tried to address a new line of research in patent analyses, namely the role of the patent attorney/representative in the filing process. The literature, especially the empirical literature, on this issue is rather limited so far, which might in part be a result of limited availability of ready to use data. In consequence, a dataset was generated for EPO patents and their representatives.

The analysis of the structure of representatives at the EPO shows a high concentration in absolute as well as in relative terms in Germany and the UK, with some activity in other larger applicant countries like France, Italy, Sweden, or the Netherlands. As a matter of fact the most experienced representatives are located in Germany and the UK, but this result might be biased due to high numbers of subsequent filings in these countries, originating mainly in the US or also Japan. Explanations for this effect are language advantages in the case of the UK (and also Ireland) and geographical proximity to the EPO as well as economies of scale in the case of Germany.

Large companies are responsible for the majority of the representatives' workload in Germany and the UK, but also in France, the Netherlands, Belgium, or Sweden—of course, mainly because many large and also multinational companies have their headquarters or IPR departments in these countries. The much lower absolute numbers of patents in southern and eastern European countries mainly stem from SMEs, individual inventors, and in some countries—for example, Portugal, Greece, Hungary, or Latvia—also stem from public research organizations or universities, which account for large shares of filings from these countries.

The multivariate analyses suggest that the (financial) resource endowment is a decisive factor in the hiring of patent attorneys. Large enterprises are able to employ the most experienced representatives, while the representatives of individual inventors are the least experienced in our dataset. While the experience has no significant impact on the probability of getting a patent granted or refused, the probability of a withdrawal increases with increasing experience of the representative. This was interpreted as experienced attorneys act more often strategically than their less experienced colleagues. In addition, it was shown that the patents of more experienced representatives were significantly less often opposed than the ones by less experienced attorneys. However, the large number of non-European filings processed in some countries might intervene, but was not separately estimated in this paper. In sum, the results show that it is possible to analyze the role of the patent attorney in the filing process, and that this is a relevant but so far under-explored dimension in the economics of patents field. The experience of the representative has a considerable impact on the outcome.

The role and characteristics of representatives on patent filings as an object of economic research is a topic that is still in its infant stage. Only limited empirical evidence is available so far, and methods and indicators are far from having been established. The analyses and discussions presented in this chapter suggest further distinctions, for example, between first and subsequent filings in future research. A more detailed analysis of internal versus external, as well as registered, European attorneys versus other legal representatives seems a reasonable distinction to be taken into account as well. A fine tuning of the experience measure also seems appropriate. We used the cumulated number of all patent filings being represented by the attorney in the past. A differentiation of the experience accumulation by different kinds of patent applications—here, we mainly argued for a distinction between subsequent PCT filings entering the regional phase versus direct filings at the EPO or at a national office in Europe—could also be worthwhile. In addition, taking into account the size of the law firm or the *organizational knowledge* of a law firm measured by the cumulated number of patents processed by the particular firm (instead of individual representative) could be another dimension to be examined. The rather rudimentary theoretical foundation with some basics of the human capital theory that was mentioned in this chapter deserves further exploration as well.

Acknowledgments. Parts of this work was funded by the European Commission under the contract number: MARKT/2014/222/B2/SE/LV

References

35.1 K. Blind, J. Edler, R. Frietsch, U. Schmoch: Motives to patent: Empirical evidence from Germany, Res. Policy **35**(5), 655–672 (2006)

35.2 European Patent Office (EPO): The European Patent Convention, Article 134, http://www.epo.org/law-practice/legal-texts/html/epc/2013/e/ar134.html (2018)

35.3 European Patent Office: How to get a European patent, guide for applicants, http://www.epo.org/applying/european/guide-for-applicants.html (2017)

35.4 J. Danguy, B. van Pottelsberghe de la Potterie: Patent fees for a sustainable EU patent system, World Pat. Inf. **33**(3), 240–247 (2011)

35.5 G. de Rassenfosse, B. van Pottelsberghe de la Potterie: Per un pugno di dollari: A first look at the price elasticity of patents, Oxf. Rev. Econ. Policy **23**(4), 588–604 (2007)

35.6 P. Hanik: Patent design using FAST, Value World **32**(3), 30–40 (2009)

35.7 W.M.H.B.U. Hussain, M.N.A. Rahman, N.I. Yaakub, Z.A. Zainul, W.K. Mujani: Online legal marketing: How lawyers assist in commercialization of university research. In: *Int. Conf. Educ. Manag. Technol., ICEMT 2010* (2010)

35.8 B.P. Nagori, V. Mathur: Basics of writing patent non-infringement and freedom-to-operate opinions, J. Intellect. Prop. Rights **14**(1), 7–13 (2009)

35.9 M. Reitzig: Improving patent valuations for management purposes—Validating new indicators by analyzing application rationales, Res. Policy **33**(6/7), 939–957 (2004)

35.10 D. Somaya, I.O. Williamson, N. Lorinkova: Gone but not lost: The different performance impacts of employee mobility between cooperators versus competitors, Acad. Manag. J. **51**(5), 936–953 (2008)

35.11 S. Macdonald, B. Lefang: Innovation and the patent attorney, Prometheus **15**(3), 329–343 (1997)

35.12 C. Li, T. Lan, S.J. Liu: Patent attorney as technology intermediary: A patent attorney-facilitated model of technology transfer in developing countries, World Pat. Inf. **43**, 62–73 (2015)

35.13 G. Buenstorf, M. Geissler: Not invented here: Technology licensing, knowledge transfer and innovation based on public research, J. Evol. Econ. **22**(3), 481–511 (2012)

35.14 K. Hussinger, A. Wastyn: In search for the not-invented-here syndrome: The role of knowledge sources and firm success, R&D Management **46**, 945–957 (2016)

35.15 R. Katz, T.J. Allen: Investigating the not invented here (NIH) syndrome: A look at the performance, tenure, and communication patterns of 50 R&D project groups, R&D Management **12**(1), 7–20 (1982)

35.16 K.S. Gaudry: The lone inventor: Low success rates and common errors associated with pro-se patent applications, PLoS One **7**(3), e33141 (2012)

35.17 C. Koller, B. Ebersberger: How do characteristics of patent attorneys influence patent quality? In: *DRUID15, Rome* (2015)

35.18 R. Frietsch, P. Neuhäusler, O. Rothengatter: Which road to take? Filing routes to the European patent office, World Pat. Inf. **35**(1), 8–19 (2013)

35.19 A. Gambardella, D. Harhoff, B. Verspagen: The value of European patents, Eur. Manag. Rev. **5**(2), 69–84 (2008)

35.20 K.A. Moore: Populism and patents, N.Y. Univ. Law Rev. **82**(1), 69–111 (2007)

35.21 M. Reitzig, S. Wagner: Patently (un) clear, Bus. Strategy Rev. **21**(1), 28–33 (2010)

35.22 D.S. Abrams, A.H. Yoon: The luck of the draw: Using random case assignment to investigate attorney ability, Univ. Chicago Law Rev. **74**(4), 1145–1177 (2007)
35.23 J. Szmer, S.W. Johnson, T.A. Sarver: Does the lawyer matter? Influencing outcomes on the supreme Court of Canada, Law Soc. Rev. **41**(2), 279–304 (2007)
35.24 G. de Rassenfosse, H. Dernis, G. Boedt: An introduction to the PATSTAT database with example queries, Aust. Econ. Rev. **47**, 395–408 (2014)
35.25 European Patent Office (EPO): PATSTAT–Backbone data set for statistical analyses, https://www.epo.org/searching-for-patents/business/patstat.html#tab-1 (2018)
35.26 R. Frietsch, P. Neuhäusler, O. Rothengatter: *SME Patenting–An Empirical Analysis in Nine Countries*, Discussion Papers Innovation Systems and Policy Analysis, Vol. 36 (Fraunhofer ISI, Karlsruhe 2013)
35.27 M. Du Plessis, B. Van Looy, X. Song, T. Magerman: *Data Production Methods for Harmonized Patent Indicators: Assignee Sector Allocation*, EUROSTAT Working Paper and Studies (EUROSTAT, Luxembourg 2009)
35.28 T. Magerman, J. Grouwels, X. Song, B. Van Looy: *Data Production Methods for Harmonized Patent Indicators: Patentee Name Harmonization*, EUROSTAT Working Paper and Studies (EUROSTAT, Luxembourg 2009)
35.29 B. Peeters, X. Song, J. Callaert, J. Grouwels, B. Van Looy: *Harmonizing Harmonized Patentee Names: An Exploratory Assessment of top Patentees*, EUROSTAT Working Paper and Studies (EUROSTAT, Luxembourg 2009)
35.30 U. Schmoch: *Concept of a Technology Classification for Country Comparisons. Final Report to the World Intellectual Property Organization (WIPO)* (Fraunhofer ISI, Karlsruhe 2008)
35.31 S. Daimer, M. Hufnagl, R. Frietsch, R. Lindner, P. Neuhäusler, O. Rothengatter: *Die Leistungsfähigkeit des deutschen Innovationssystems – Eine Bilanz der Patentaktivitäten nach zehn Jahren Hightech-Strategie*, Fraunhofer ISI Discussion Papers Innovation Systems and Policy Analysis, Vol. 57 (Fraunhofer Institute for Systems and Innovation Research ISI, Karlsruhe 2018)
35.32 P. Neuhäusler, O. Rothengatter, R. Frietsch: Patent applications–Structures, trends and recent developments 2017. In: *Studien zum deutschen Innovationssystemm Berlin 2015*, Vol. 4, ed. by Expertenkommission Forschung Innovation (EFI, Berlin 2018)
35.33 P. Neuhäusler, O. Rothengatter, R. Frietsch: Patent applications–Structures, trends and recent developments 2016. In: *Studien zum deutschen Innovationssystem*, Vol. 4, ed. by Expertenkommission Forschung Innovation (EFI, Berlin 2017)
35.34 G.S. Becker: *The Economic Approach to Human Behavior* (Univ. of Chicago Press, Chicago 1990)
35.35 G.S. Becker, B.R. Chiswick: Education and the distribution of earnings, Am. Econ. Rev. **56**(1/2), 358–369 (1966)
35.36 J. Mincer: Investment in human capital and personal income distribution, J. Polit. Econ. **66**, 281–302 (1958)
35.37 J. Mincer: *Schooling, Experience, and Earnings, Human Behavior and Social Institutions*, National Bureau of Economic Research (Columbia Univ. Press, New York 1974)
35.38 J. Mincer: Human capital and earnings. In: *Wealth, Income, and Inequality*, ed. by A.B. Atkinson (Oxford Univ. Press, Oxford 1980) pp. 103–128
35.39 T.W. Schultz: Investment in human capital, Am. Econ. Rev. **51**(1), 1–17 (1961)
35.40 T.W. Schultz: On investing in specialized human capital to attain increasing returns. In: *The State of Development Economics. Progress and Perspectives*, ed. by G. Ranis, T.P. Schultz (Basil Blackwell, Oxford 1988) pp. 339–352
35.41 J. Heckman, L. Lochner, P. Todd: *Fifty Years of Mincer Earnings Regressions*, NBER Working Paper Series, Vol. 9732 (National Bureau of Economic Research, Cambridge 2003)
35.42 J.J. Heckmann: *The Economics, Technology and Neuroscience of Human Capability Formation*, NBER Working Paper Series (National Bureau of Economic Research, Cambridge 2007)

Rainer Frietsch

Competence Center Policy – Industry – Innovation
Fraunhofer Institute for Systems and Innovation Research ISI
Karlsruhe, Germany
rainer.frietsch@isi.fraunhofer.de

Rainer Frietsch is Deputy Head of the Competence Center Policy - Industry - Innovation and Coordinator of the Business Unit Innovation Indicators at the Fraunhofer Institute for Systems and Innovation Research ISI. Rainer completed business training in wholesale and import/export trade. After that he studied Social Science at the University of Mannheim. He holds a PhD in Economics from the Karlsruhe Institute of Technology (KIT).

Peter Neuhäusler

Competence Center Policy – Industry – Innovation
Fraunhofer Institute for Systems and Innovation Research ISI
Karlsruhe, Germany
peter.neuhaeusler@isi.fraunhofer.de

Peter Neuhäusler is a Project Manager at the Competence Center Policy - Industry - Innovation at the Fraunhofer Institute for Systems and Innovation Research (Karlsruhe). He studied Social Sciences, majoring in Micro- and Macro-Sociology, Applied Social Research, and Quantitative Methodology at the University of Mannheim and graduated with a diploma in 2008. Peter Neuhäusler was awarded his Doctoral degree in Innovation Economics at the Technical University of Berlin in 2012.

36. Exploiting Images for Patent Search

Ilias Gialampoukidis, Anastasia Moumtzidou, Stefanos Vrochidis, Ioannis Kompatsiaris

Patent offices worldwide receive considerable numbers of patent documents that aim at describing and protecting innovative artifacts, processes, algorithms, and other inventions. These documents apart from the main text description may contain figures, drawings, and diagrams in an effort to better explain the patented object. Two main directions are presented in this chapter; concept-based and content-based patent retrieval. Concept-based search utilizes textual and visual information, fusing them in a classification late fusion stage. Conversely, content-based retrieval is based on the shape/content information from patent images and is therefore based on the visual descriptors that are extracted from binary images. Concepts are extracted using classification techniques, such as support vector machines and random forests. Adaptive hierarchical density histograms serve as binary image retrieval techniques that combine high efficiency and effectiveness, while being compact and therefore capable of dealing with large binary image databases. Given the vast number of images included in patent documents, it is highly significant for the patent experts to be able to examine them in their attempt to understand the patent contents and identify relevant inventions. Therefore, patent experts would benefit greatly from a tool that supports efficient patent image retrieval and extends standard figure browsing and metadata-based retrieval by providing content-based search according to the query-by-example paradigm.

36.1 How Patent Document Analysis Evolved

The number of patent documents describing and protecting innovative artifacts, processes, and other inventions that are submitted to patent offices worldwide is growing steadily, thus necessitating the development of advanced patent search technologies that take into consideration that complexity and the unique features of patents. In this direction, the intellectual property and the information retrieval communities have shown great interest in patent image search, which is expressed with common research activities in relevant conferences (e. g., the *International Retrieval Facility Conference* (IRFC) and *The Conference and Labs of*

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_36

the Evaluation Forum Track on Intellectual Property (CLEF-IP). Although a typical patent document contains multimedia information, the majority of the search engines rely upon text to provide search functionalities so the ideas to be patented are usually described in text format. However, many patents include a drawing section that may contain figures, drawings, diagrams as a means for further describing patented inventions and thus they should be considered in the patent retrieval systems given that images by nature are not affected by the applicant's language and remain intact despite the evolvement of the scientific terminology over the years. As a consequence, an increasing interest in patent image and multimodal search has been shown by the intellectual property and the information retrieval communities. The retrieval functionalities of patent image and multimodal systems should extend beyond figure browsing and metadata-based retrieval to include content-based searches including query-by-example search, image classification, and concept extraction.

This chapter analyzes techniques from document image analysis that are employed for patent preprocessing in order to extract the patent images and retrieve visually similar images using content-based and concept-based techniques. Regarding content-based search, it will discuss applications of image processing and visual feature extraction algorithms such as salient-based (e. g., scale invariant features transform (SIFT) [36.1], speeded-up robust features (SURF) [36.2]) as well as color layout and edge histograms, that capture the global characteristics of images (e. g., colored images, images depicting objects, images with landscapes). Moreover, the chapter presents methods oriented towards capturing features of binary images such as the adaptive hierarchical density histogram (AHDH). Retrieval is performed based on similarity metrics using such features, in an effort to imitate the way humans perceive visual similarity.

Following the trend of modern image retrieval approaches that steadily move towards concept-based image search, we present concept extraction approaches for patent images. Concept extraction techniques involve identifying common characteristics among images that classify them into a specific semantic category or depict a specific concept. Specifically, it involves extracting low-level features from patent images combined with supervised machine learning algorithms, such as support vector machines and random forests (RFs). Aiming to boost the effectiveness in concept extraction, hybrid approaches are applied that consider both textual and visual features.

Bridging the semantic gap in patent document search has also been discussed in [36.3]. Firstly, a text/graphic separation module separates the textual elements from the graphical ones, through an optical character recognition (OCR) engine on the text layer, while the nodes and edges on the graphical layer are analyzed. The vectorial-based approach requires a conversion module that transforms the raw pixel image into a vectorial representation. Query expansion methods have also been used in patent image retrieval [36.4], where *LambdaRank* was employed to improve patent retrieval performance by combining different query expansion methods with different text fields weighting strategies of different resources. Patent query logs have also been exploited in [36.5] to expand the query terms that are used by patent examiners. Patent retrieval systems are also compared in the short review of [36.6], but the evaluation of a patent retrieval system is a challenging task [36.7].

The chapter discusses the results for some indicative patent image feature extraction methods [36.8] that are applied for content-based image retrieval (Sect. 36.4). Concept-based patent retrieval methods are discussed in Sect. 36.5, exploiting the involved visual information, in combination with textual information [36.9–11]. Furthermore, specific use cases and interaction modes are presented [36.12], as introduced in the search requirements (Sect. 36.2), demonstrating the benefits and impact of such technologies to patent examiners and industrial developments. Finally, we discuss limitations from recent research and future directions (Sect. 36.6).

36.2 Patent Search Scenario and Requirements

An example of a mechanical search as described by a professional patent searcher [36.13] is the following: "A dancing shoe with a rotatable heel to allow rapid pivoting about your heel. In a preferred embodiment, the heel should have ball bearings." The patent searcher has to distill the core for this disclosure, which could be used as the basis of the upcoming search. Therefore, for the particular example, the core could be expressed by the following concepts:

- Concept 1: *Dancing shoe*
- Concept 2: *Rotating heel*
- Refined Concept 2: *Rotating heel with ball bearings*.

Afterwards, the patent searcher proceeds by keyword-based search on the aforementioned concepts and classification areas. In many cases, such as the aforementioned example, both the important information and the core are described using figures. Thus, it is evident that the existence of such concepts that describe the figures inside patents (i. e., dancing shoe and rotating heel for this example) would simplify and expedite the search of the patent expert.

To that end, it is required to support concept extraction from images, through the exploitation of the heterogeneous information (textual and visual) per patent (image and figure description). However, the image description is not sufficient for the concept extraction in patent images, since many figures can be associated with misleading or incomplete descriptions (e. g., references to other figures or parts of the patent). Moreover, the automatic mapping of figure captions to the figures themselves is not a trivial task and it can be further complicated in the case of handwritten figure labels that cannot be automatically recognized as labels. Another reason is that many patents (and consequently the included image descriptions) are written in languages that present challenges in translation (e. g., Chinese), and thus patent search can only rely on visual information.

Moreover, a set of requirements regarding the system performance has been reported in [36.10]. Firstly, the system should be on the one hand scalable considering that it needs to cope with vast amounts of content (in the order of millions of patent images), to avoid expensive fusion algorithms, but on the other hand the representations of the patent images should suffice so that the concept detectors can perform satisfactory, i. e., accuracy should be around 85–90% (depending also on the concept characteristics). Secondly, a significant number of concepts that are indicative of the patent images of each international patent classification (IPC) class and subclass have to be defined by patent experts, while manual annotation of relevant examples is required for training the machine learning algorithms, and thirdly, the framework needs to use open technologies and standards, in order to be easily adaptable to the established patent search platforms.

36.3 Feature Extraction

The extraction of features involves a preprocessing stage before generating visual and textual features. The initial document processing stage is required in order to extract all the desired images and subfigures embedded in the document, along with related metadata. The decomposition of each document page of the drawing section of the patent PDF file and the extraction of the figures found in the retrieved pages can be performed using automatic segmentation techniques [36.14].

In the sequel, each image is served into the feature extraction component, where the visual and textual based features are produced.

36.3.1 Visual Features

The extraction of global concepts requires the use of visual features that represent the image in a global manner and which can handle the special features and complexity of patent images. The distinctive characteristic of patent figures is that they are usually black-and-white and that they depict technical information in diagrammatic form. As far as coloring is concerned, until recently all the patent figures included in the patent documents were black-and-white. A few Korean patent publications deviate from the general rule and include color images. Most visual representation features are based on color and texture, which are absent in most patent images, thus we need to apply an algorithm that considers other visual elements, such as geometry and pixel distribution. To this end, we apply the algorithm proposed in [36.8] to extract the adaptive hierarchical density histograms (AHDHs) as visual feature vectors. The AHDH was devised specifically to deal with binary and complex images and the overall feature extraction method can be found in Sect. 36.4.1. The experiments realized using several patent datasets revealed that the AHDH outperforms the other state-of-the-art methods, as presented in [36.8].

36.3.2 Textual Features

In order to exploit the textual descriptions related to each figure in the patent document, the figure captions are processed and textual features are extracted from the patent images. In particular, a bag-of-words (BoW) approach is applied, modeling each figure with a vector. The BoW representation method is a simplifying assumption used in natural language processing and information retrieval, where a text description such as a sentence or a document is represented as an unordered collection of words, without considering grammar and word order. The generation of such a vector requires the definition in advance of a vocabulary that consists of the most frequently used words of this dataset. Then for

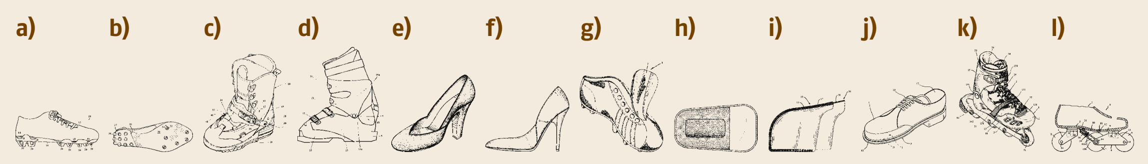


Fig. 36.1a–l Sample of patent images and its associated concepts and description; **(a,b)** cleat, **(c,d)** ski boot, **(e,f)** high heel, **(g,h)** tongue, **(i,j)** toe caps, **(k,l)** roller skate. Reprinted from [36.10], with permission from Elsevier

each word, the term frequency is computed, weighted by inverse document frequency scores (tf-idf) in order to minimize the significance of words appearing too often

$$\text{tf-idf}_{ij} = \frac{n_{id}}{n_d} \log \frac{D}{n_i}\,, \tag{36.1}$$

where n_{id} is the number of occurrences of word i in document d, n_d is the number of words in document d, n_i is the number of occurrences of word i in the whole database, and D is the total number of documents in the database. Apparently, an initial preprocessing on the raw text is required for the removal of stop-words and for stemming the remaining words, using for example Porter's stemming algorithm [36.15].

Several concepts appear in Fig. 36.1; Cleat: A short piece of rubber, metal et cetera, attached to the bottom of a sports shoe used mainly for preventing someone from slipping (a), (b); Ski boot: A specially made boot that fastens onto a ski (c), (d); High heel: Shoes with high heels (e), (f); Tongue: The part of a shoe that lies on top of your foot, under the part where you tie it (g), (h); Toe caps: A reinforced covering of leather or metal for the toe of a shoe or boot (i), (j); Roller skate: A shoe or boot with two or four wheels or casters attached to its sole for skating on hard surfaces (k), (l). All images are manually extracted from around 300 patents belonging to A43B and A63C IPC subclasses, which contain parts of footwear.

36.4 Content-Based Patent Image Retrieval

Content-based image retrieval (CBIR) [36.16] has become one of the most developed fields of computer vision for image retrieval [36.17]. CBIR systems use different types of queries including a query-by-example image, sketch, or region and return relevant images from a given database, by considering textual annotation, media metadata of low-level visual features. CBIR techniques can be applied in several domains, either as standalone implementations supporting queries-by-example, or as complementary modules of an integrated framework that incorporates several retrieval options, such as text and concept search (Sect. 36.5), in order to improve and enhance the results.

CBIR algorithms represent image content with representative feature vectors that employ color, texture, and shape characteristics [36.18]. Although modern technologies generate high-quality color images, there are still many image databases that contain exclusively binary (i. e., black-and-white) images, such as patent images, trademarks, technical drawings, or other specific applications like road signs [36.19], botanical collections [36.20], or medical images [36.21]. The pixels of such images are generated from original black and white documents, which are scanned to gray-scale ones, and then thresholded to the final binary images. Therefore, they contain no color and minimum texture information. As a result, in order to retrieve effectively such images, a content-based binary image analysis technique that considers the image geometric information accurately should be applied [36.22]. The MPEG-7 standard has summarized the required criteria for techniques that measure the similarity between binary images [36.22]: high discrimination capability, invariance to geometric transformations, i. e., rotation, translation and scaling, computational efficiency, robustness to distortion and noise, compactness, generality of the application, and handling large image databases without heavy performance degradation.

However, the aforementioned requirements pose several challenges to binary image retrieval given that many binary images, including technical drawings and patent images, are produced from noisy analog sketches that are exposed both to arbitrary time degradation and content degrading during digitizing. Moreover, it is often the case that binary images are sketches created with a drawing style that is determined exclusively by the creator's preferences and the specific application.

The following section presents the adaptive hierarchical density histogram (AHDH) which employs an adaptive pyramidal decomposition of the image into regions based on the recursive calculation of geometric centroids, and produces a robust binary image descriptor by generating the density histogram of each region.

36.4.1 Adaptive Hierarchical Density Histogram (AHDH)

The algorithm for the generation of AHDH involves two main parts: (a) the region partitioning based on the generation of the adaptive geometric centroids as proposed in [36.23] and (b) the adaptive hierarchical density histogram generation as proposed in [36.8]. Figure 36.2 depicts a schematic view of the algorithm, while Table 36.1 contains the definitions of the most important variables involved in the theoretical analysis of the algorithm.

Binary images are treated as two-dimensional (2-D) planes where the set of shape pixels B is defined. The number of pixels that belong to the shape are defined as N. Before initializing the algorithm, we normalize the pixel coordinates in order to be translation-invariant. Moreover, we define as R_i^l, $i = \{1, 2, \dots, 4^{(l-1)}\}$ the i-th rectangular region of a binary image, which corresponds to the set of black pixels included in the region. Regions are produced progressively, in an iterative manner, by using the following adaptive geometrical centroid estimation algorithm, which leads to the construction of an adaptive asymmetric orthogonal grid that covers the entire (black-and-white) image. Since region partitioning is triggered by an iterative algorithm, l is the number of iteration, or *level*. Each rectangular region R_i^l consists of two main features: first, the number of black pixels N_i^l that lie in it; and second, its area E_i^l.

Table 36.1 Necessary notation

Symbol	Definition
B	The set of black pixels that belong to a binary image
N	The amount of black pixels that belong to a binary image
R_i^l	The i-th region of the l-th level of a binary image
N_i^l	The number of black pixels that belong to the region R_i^l
E_i^l	The area of a region R_i^l
$SR_{i,j}^l$	The j-th subregion of the region R_i^l
$N_{i,j}^l$	The number of black pixels that belong to the region R_i^l
$E_{i,j}^l$	The area of subregion $SR_{i,j}^l$
$d_{i,j}^l$	The density of a subregion $SR_{i,j}^l$ of a region R_i^l
$\hat{d}_{i,j}^l$	The relative density of a subregion $SR_{i,j}^l$ of a region R_i^l
l_{d}:	The minimum level for which a quantized feature is used
w_i	The i-th distribution word that is used for quantization of $\hat{d}_{i,j}^l$
$h(w_i)^l$	The histogram value of the i-th distribution word in the l-th level
L	The lexicon that is formed from all distribution words
FV_l	The density feature of the l-th level that was constructed by the employment of density
$\hat{FV}_l$	The relative density feature of the l-th level that was constructed by the employment of relative density

For each iteration or level l, a region partitioning procedure is applied which involves estimating the geometric centroid of all regions R_i^l produced and the splitting of each region into four subregions $SR_{i,j}^l$, $j =$

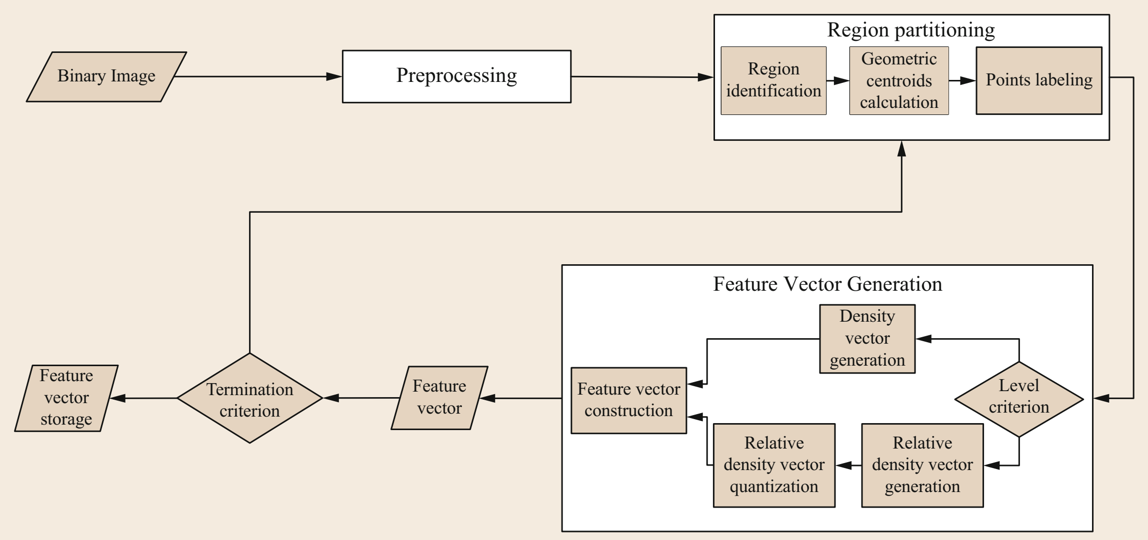


Fig. 36.2 Block diagram for the generation of the adaptive hierarchical density histogram (reprinted from [36.8], with permission from Elsevier)

$\{1, 2, 3, 4\}$ using the geometric centroid as the center. The initial region (i. e., for $l = 1$) is the whole image and for $l > 1$ all regions of level l are the subregions of level $l - 1$. This iterative procedure is terminated when the termination level is reached.

The feature extraction procedure depends on the level value; in *lower* levels, the feature is a vector of the distribution of the N_i^l black pixels into the four subregions, while in *higher* levels a two-class classification of each subregion is employed. The two classes are labeled as *Full* and *Empty* and are defined by the percentage of N_i^l black pixels that lie in the subregion in comparison to the percentage of the region's area E_i^l that belongs to the subregion. The combination of the classes of all four subregions of a region produces a *distribution word*. This *distribution word* histogram is used as the levels feature vector.

In the following section, we will provide some insight regarding the rationale behind the selection of different feature extraction procedures for different level values.

Region Partitioning

Region partitioning is performed using the geometric centroid extraction method that was introduced in [36.24]. Thus, in each level l, a number of 4^{l-1} regions R_i^l has to be processed. We consider a separate Cartesian coordinates system for each region R_i^l, and the equation that captures the geometrical centroid [36.24] of each nonempty region R_i^l is the following

$$x_c = \frac{\sum_{(x,y)\in B_i^l} x}{N_i^l}, \qquad y_c = \frac{\sum_{(x,y)\in B_i^l} y}{N_i^l}. \tag{36.2}$$

N_i^l denotes the amount of black-pixels set B_i^l in the processed region R_i^l, and (x, y) are the pixel coordinates. These centroids partition the image plane into an adaptive hierarchical biased orthogonal grid (Fig. 36.3).

Adaptive Hierarchical Density Histogram Generation

Two different approaches are considered for the construction of the feature vector: (a) density features and (b) quantized relative density features. For levels l less than an experimentally defined level l_d, density features are estimated, while if $l \geq l_d$ quantized relative density features are computed. The features are calculated for all regions at each level and the overall feature vector is updated at each iteration of the algorithm.

Thus, in the l-th level (i. e., after l iterations), 4^{l-1} regions R_i^l are identified, with area E_i^l and N_i^l black pixels. For each of these regions, the centroid estimation results in a partition into four new subregions $SR_{i,j}^l$, $j = \{1, 2, 3, 4\}$ with area $E_{i,j}^l$ and $N_{i,j}^l$ black pixels, where

$$\sum_{j=1}^{4} N_{i,j}^l = N_i^l$$

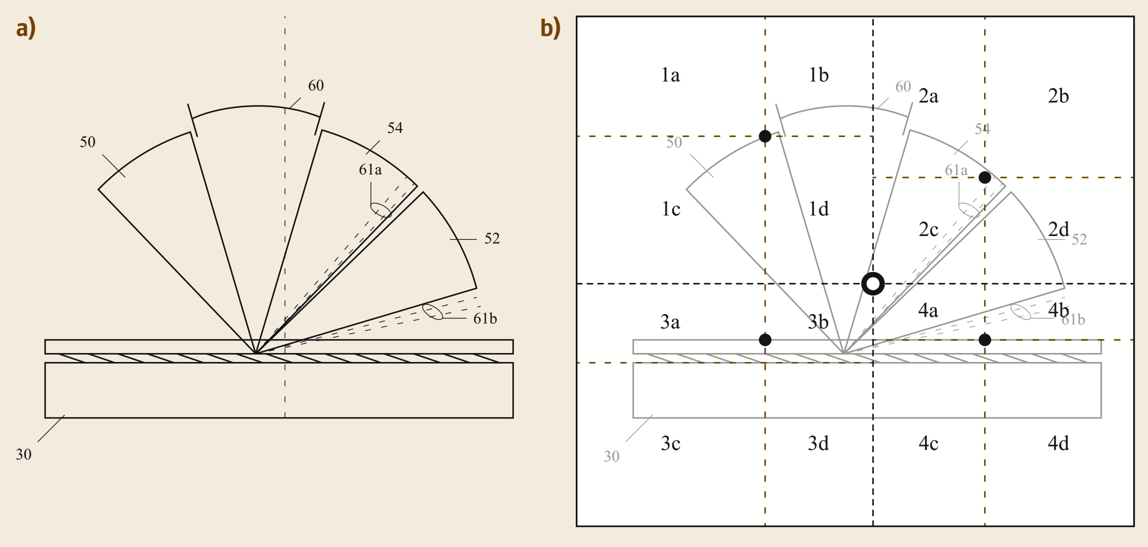


Fig. 36.3 **(a)** Binary figure 15c of patent EP 1 555 563 A2 of the European Patent Office (reprinted from [36.9], with permission from Elsevier). **(b)** First- and second-level centroids and partitions (reprinted from [36.8], with permission from Elsevier). Each subregion is characterized by a number that labels the first-level partition followed by a letter that marks the second-level partitions. It is evident that the iterative geometric centroid procedure generates an adaptive hierarchical biased orthogonal partition

and

$$\sum_{j=1}^{4} E^l_{i,j} = E^l_i \,.$$

Two variables are defined which are the density $d^l_{i,j}$ of a subregion $\mathrm{SR}^l_{i,j}$ of a region R^l_i and the relative density $\hat{d}^l_{i,j}$. Density $d^l_{i,j}$ is defined as the amount of black pixels of a subregion $\mathrm{SR}^l_{i,j}$ divided by N^l_i, while relative density $\hat{d}^l_{i,j}$ as the subregion density compared to the ratio of the subregion area $E^l_{i,j}$ over the region area E^l_i

$$d^l_{i,j} = \frac{N^l_{i,j}}{N^l_i}, \qquad \hat{d}^l_{i,j} = \frac{E^l_i N^l_{i,j}}{N^l_i E^l_{i,j}} \,, \tag{36.3}$$

where $1 \le i \le 4^{l-1}$ and $1 \le j \le 4$. Subsequently a $4^{l-1} \times 4$ feature array FA_l is constructed by the densities of each new subregion.

$$\mathrm{FA}_l = \begin{vmatrix} d^l_{1,1} & d^l_{1,2} & d^l_{1,3} & d^l_{1,4} \\ d^l_{2,1} & d^l_{2,2} & d^l_{2,3} & d^l_{2,4} \\ \cdots & \cdots & \cdots & \cdots \\ d^l_{4^{l-1},1} & d^l_{4^{l-1},2} & d^l_{4^{l-1},3} & d^l_{4^{l-1},4} \end{vmatrix} \,. \tag{36.4}$$

Finally, the FA_l is serialized to form the l-level feature vector FV_l

$$\begin{aligned} \mathrm{FV}_l = \Big\{ & d^l_{1,1}, d^l_{1,2}, d^l_{1,3}, d^l_{1,4}, d^l_{2,1}, d^l_{2,2}, d^l_{2,3}, d^l_{2,4}, \\ & \ldots, d^l_{4^{l-1},1}, d^l_{4^{l-1},2}, d^l_{4^{l-1},3}, d^l_{4^{l-1},4} \Big\} \,. \end{aligned} \tag{36.5}$$

In the same way, the feature vector $\hat{FV}_l$ is constructed, which involves $\hat{d}^l_{i,j}$ instead of $d^l_{i,j}$

$$\begin{aligned} \hat{FV}_l = \Big\{ & \hat{d}^l_{1,1}, \hat{d}^l_{1,2}, \hat{d}^l_{1,3}, \hat{d}^l_{1,4}, \hat{d}^l_{2,1}, \hat{d}^l_{2,2}, \hat{d}^l_{2,3}, \hat{d}^l_{2,4}, \\ & \ldots, \hat{d}^l_{4^{l-1},1}, \hat{d}^l_{4^{l-1},2}, \hat{d}^l_{4^{l-1},3}, \hat{d}^l_{4^{l-1},4} \Big\} \,. \end{aligned} \tag{36.6}$$

The employment of the density or relative density vector is determined by the partition level. Specifically, experiments have shown that the density vector performs best in lower levels. Hence, the feature vector of these levels is simply the serialized density vector of (36.5). Unfortunately, the dimension of the density feature increases exponentially with the number of levels and thus it is not considered appropriate for representation in higher levels. Conversely, in the higher level an image is decomposed into exponentially smaller parts and therefore the information that each feature element contains is rapidly decreasing. In order to overcome those difficulties, we employ a simple density lexicon scheme that reduces descriptor size even for high levels. This quantization process can be applied in either density or relative density vectors (36.6).

Relative Density Vector Quantization

Contrary to centroid vectors, density and relative density vectors may also be (vector) quantized. Thus [36.8] perform vector quantization in relative density vectors.

Subregions $\mathrm{SR}^l_{i,j}$ are labeled as *Full* or *Empty*. Specifically, when

$$\hat{d}^l_{i,j} \ge 1 \,,$$

the subregion is labeled as *Full*, otherwise it is labeled as *Empty*. The motivation behind the use of the Empty and Full labels can be statistically explained. Specifically, when the black pixels of a region R^l_i fall in the four subregions following a uniform random distribution, the expected value of relative density is 1 for every subregion. Consequently, a subregion is labeled *Full* or *Empty* depending on the amount of pixels that lie in its interior compared to the statistically expected value. It should be noted that density vector quantization can not utilize such a statistical measure and the assignment of labels *Full* and *Empty* is fully arbitrary since quantization is controlled by a manually chosen between-classes threshold.

At this stage, a lexicon L of *distribution words* w is defined, which represents the 16 combinations of the 4 Full or Empty subregions of a processed region. Assuming that E corresponds to Empty and F to Full, the lexicon L has the following format

$$\begin{aligned} L &= [EEEE, EEEF, EEFE, \ldots, FFFF] \\ &= [w_0, w_1, w_2, \ldots, w_{15}] \,. \end{aligned} \tag{36.7}$$

However, the valid words that are used are actually 15. *EEEE* or w_0 is a nonvalid word, since, by the definition of relative density the amount of black pixels of a subregion can not be less than the expected value for all four subregions of any region. Therefore, according to the above lexicon, the quantized feature for each level is

$$\hat{FV}_{\mathrm{q},l} = [h(w_1)^l, h(w_2)^l, \ldots, h(w_{15})^l] \,, \tag{36.8}$$

where $h(w_i)^l$ is the normalized histogram, estimated by counting the number of appearances of the respective word, normalized over the total number of subregions in the level l (i. e., $h(w_1)^5$ is the histogram value of the word EEEF in the 5 th level).

Eventually, the newly constructed feature vector is merged with the feature vector that was generated during the previous iterations

$$FV = \left[\mathrm{FV}_1\, \mathrm{FV}_2\, \ldots \mathrm{FV}_{l_\mathrm{d}-1}\, \hat{\mathrm{FV}}_{\mathrm{q},l_\mathrm{d}}\, \hat{\mathrm{FV}}_{\mathrm{q},l_\mathrm{d}+1}\, \ldots \hat{\mathrm{FV}}_{\mathrm{q},l}\right],$$

where l_d, as discussed, is the first level for which quantized relative density features are extracted. When the algorithm is complete, the final FV represents the adaptive hierarchical density histogram of the image and can be used for retrieval purposes.

At this point, we should note that the relative density vectors are employed and quantized only to allow the modeler to reach the deepest levels, where local information about the black pixels topological structure exists. In the upper levels, this is both unnecessary and ineffective, given that the global distribution of black pixels is much better represented by the density vector instead of the knowledge that its relative density vector belongs to a certain class.

36.4.2 AHDH Evaluation in Patent Image Retrieval

A patent image database that includes 2000 binary images has been created in order to evaluate and compare the AHDH method against other methods used in similar cases, including edge orientation autocorrelogram (EOAC)[36.25], triple adjacent segments (tAS) [36.26], and Yang's centroid vector [36.24]. As far as the patent image database is concerned, it contains images extracted from patent documents filed to the European Patent Office [36.27]. These images are usually very complex and cannot be easily segmented into simple shapes, or into connected components with a single contour. Therefore, the contour-based descriptors of [36.25, 28], as well as the technique introduced in [36.29] that requires image partition into filled in regions, are excluded from this comparison. The images used as queries are 120 images randomly selected from the initial database and the number of relevant images found in the database for each query varies from 2 to 73. The AHDH method was tested with $l_\mathrm{d} = 3$ and $l = 10$, while all other methods were implemented with the parameters defined by the corresponding publications.

The precision-recall curves are shown in Fig. 36.4. After a careful observation, we can deduce that for identical recall rates, the AHDH precision rate is at least 20–40% higher than the other techniques, while for identical precision rates, the recall rate is 5–25% better. Table 36.2 contains the maximum F-Score performance values for all methods. Finally, we should note that when recall and precision values are equal (and consequently the F-Score is equal to them), the AHDH method leads to a mutual enhancement of at least 12.7% both for recall and precision values. The above com-

Table 36.2 Performance comparison of AHDH, EOAC [36.25], tAS [36.26], and the centroid vector method proposed in [36.24] for binary image retrieval in a patent image database

Patent image performance comparison			
Method	Recall (%)	Precision (%)	F-Score
AHDH	71.39	67.74	70.13
EOAC	62.97	52.53	59.06
Centroid vector (Yang)	49.82	59.49	52.67
tAS	48.6	53.41	50.1

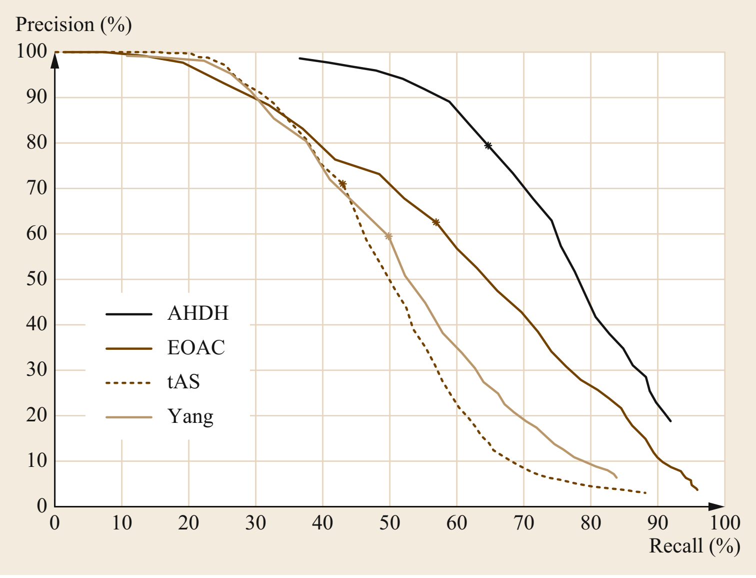


Fig. 36.4 Performance in patent binary images in a precision-recall curve

parison indicates the improved performance of AHDH over the other prominent binary image descriptors.

A query-based experiment has also been conducted [36.8], in which each image example was associated with the 25 most similar images retrieved. For this experiment, 86.9% of the manually annotated near-replicas or similar images were successfully retrieved. Figure 36.5 depicts the query image and the first retrieved images for two cases that involve visual search for technical drawings of circular shape and flowcharts.

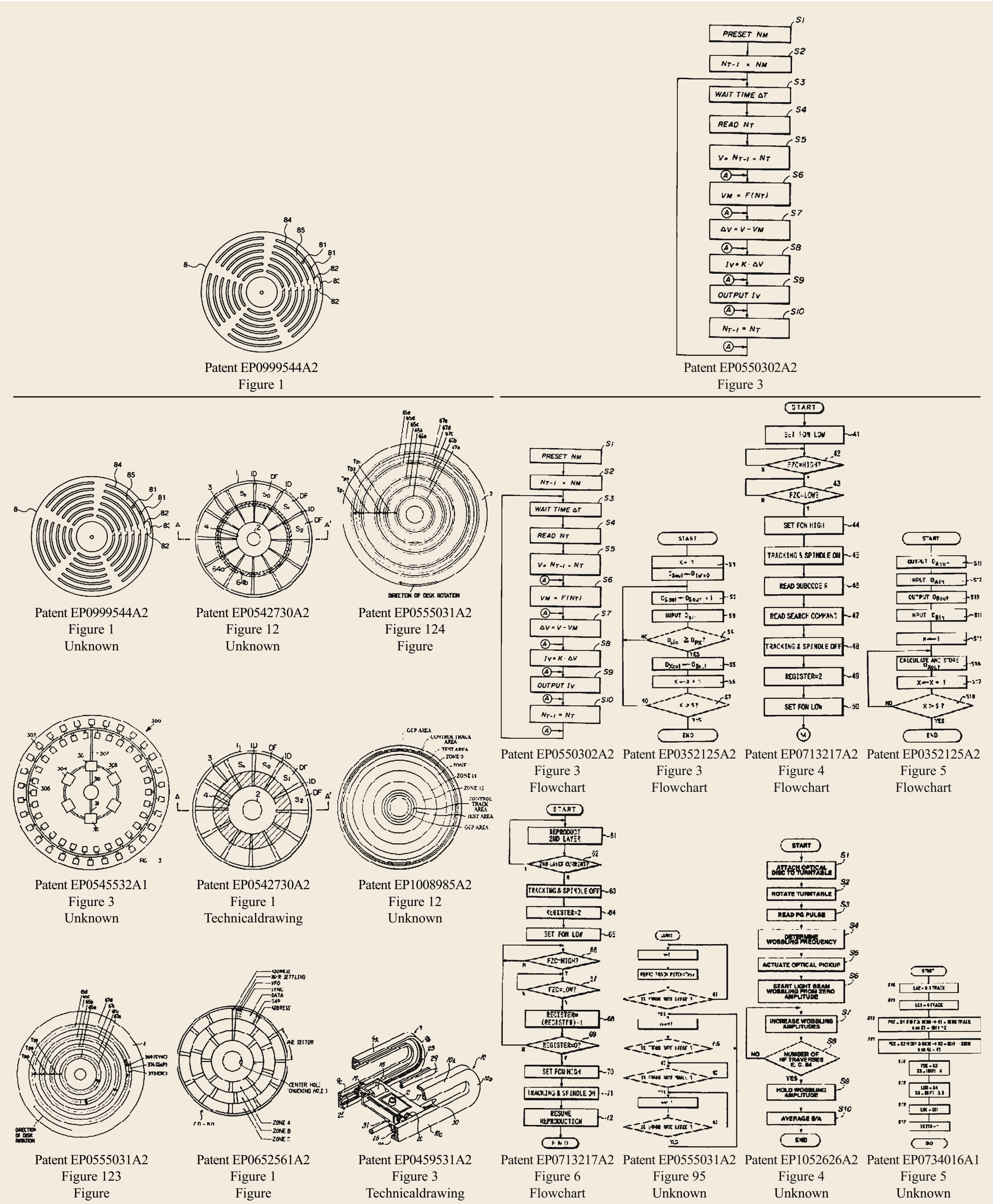


Fig. 36.5 A patent image example and the first nine retrieved similar images for flowcharts and technical drawings

36.5 Concept-Based Patent Image Retrieval

Advanced patent search tools exploit nontextual elements that play a key role in patent search through patent image retrieval. Following the latest challenges in image analysis (semantic indexing, semantic gap), patent image search deals with patent image classification [36.30] and concept extraction [36.10].

The motivation behind the interest in patent concept-based search is revealed by the scenario described earlier in Sect. 36.2, where a patent searcher searches for a dancing shoe that incorporates a rotating heel with ball bearings. The main concepts of the search are recognized by the patent searcher and they may be used as keywords and relevant classification areas. However, it is rather common that the important information is described with figures. In order to deal with such cases, it would be very helpful if the patent searcher could directly retrieve patents that contain figures depicting these concepts. Thus, the integration of concept-based retrieval functionalities in patent search systems would facilitate significantly the tasks of patent searchers.

To address the requirements of Sect. 36.2, a supervised machine learning-based framework has been proposed [36.10, 31], which is based on both visual and textual information [36.30], as illustrated in Fig. 36.6. Specifically, the procedure begins by extracting the patent images and the associated captions. In the sequel, the images and their captions are fed into the feature extraction component, where visual and textual features are obtained. Given that the framework described is based on supervised machine learning techniques, it is essential to have a training dataset for developing the models and a test dataset for evaluating the performance. Thus, the dataset used is manually annotated and split into a training and a test set. It should be noted that when splitting the image/text dataset, the images/text belonging to the same patent are kept together. Then, we train three models for each concept using:

(a) Visual features
(b) Textual features
(c) The results of the previous models (a) and (b).

The latter (c) are used as features to train a hybrid classification model for final results. After observing

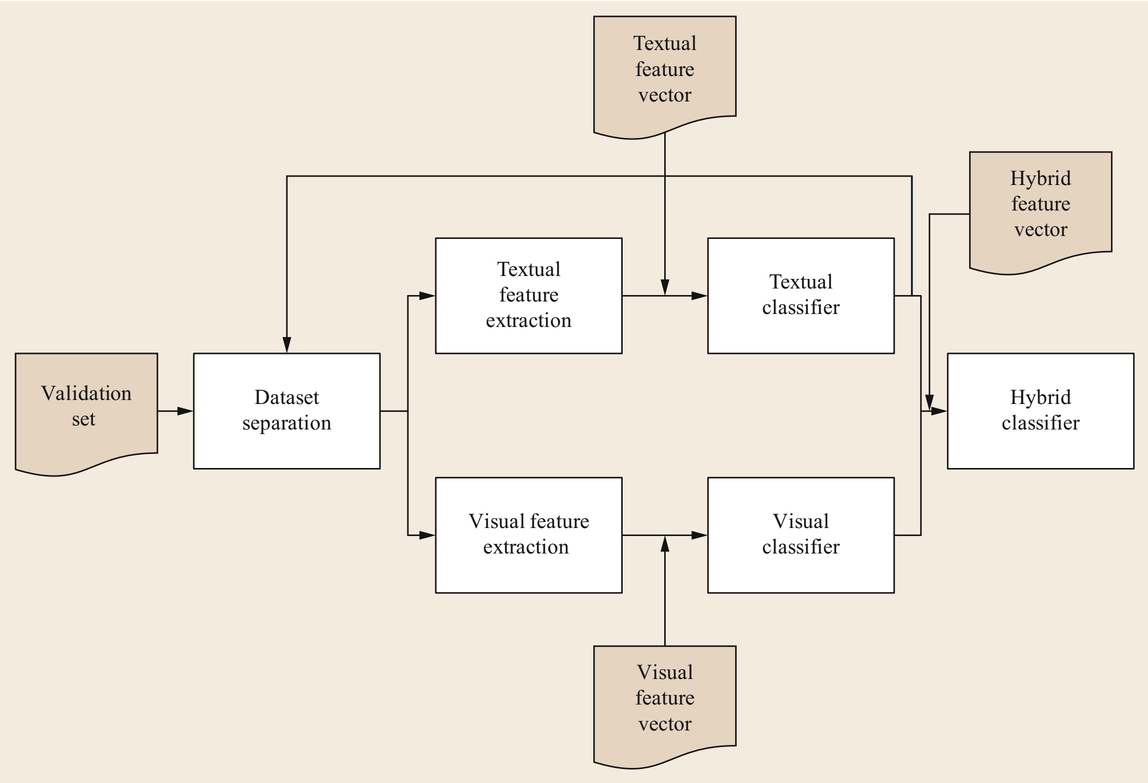


Fig. 36.6 Testing procedure for patent concept extraction to assist the *search-by-concept* functionality, as presented in [36.31]

the figure captions carefully, we discover that a significant number of captions do not provide any description of the figure, but they simply mention that the specific figure depicts another view of another figure (e. g., Fig. 36.2 is the front view of Fig. 36.1). This kind of description is one of the main reasons for retrieving low-quality results in text-based concept extraction [36.10]. Therefore in order to limit the negative impact of such captions, we further split the test set into two sets (i. e., A and B) based on the following rule: the figures that have a description that points to another figure of the same patent belong to set B, while the rest belong to set A. Then, during the testing phase (Fig. 36.6), the figures along with their captions that belong to set A are fed into the textual and visual feature extraction components and their features are used as input to the textual and visual classifiers, respectively. Afterwards, the hybrid classifier provides the final confidence score for each concept. Regarding the images contained in set B, the missing parts of their descriptions are replaced by the results of the textual classifier in a recursive way and this process continues recursively until all the captions of the figures in set B are updated.

36.5.1 Classification Techniques for Concept Extraction in Patent Images

Concept-Extraction Using Support Vector Machines (SVM)

Support vector machines (SVM) are supervised learning models that analyze data and recognize patterns; they have been successfully applied to several classification and regression problems. Specifically, SVM is a discriminative classifier that is formally defined by separating decision hyperplanes. These hyperplanes are constructed in a multidimensional space and aim at separating cases of different class labels.

A simple schematic example is depicted in Fig. 36.7, where the objects belong either to class green stars or red circles. The separating line (i. e., decision hyperplane) defines a boundary on the right side of which all objects are green and to the left of which all objects are red. Thus, any new object that falls to the right is labeled/classified as green, otherwise as red.

At the supervised learning approach of the framework depicted in Fig. 36.6, classification is applied both on the textual and visual features, presented in Sect. 36.3, that describe the patent images. However, other classification techniques have shown that [36.32] can further improve the retrieved results of a patent search scenario. In the following, we present another framework (Fig. 36.8), which also combines textual and visual information from patents, that is able to detect

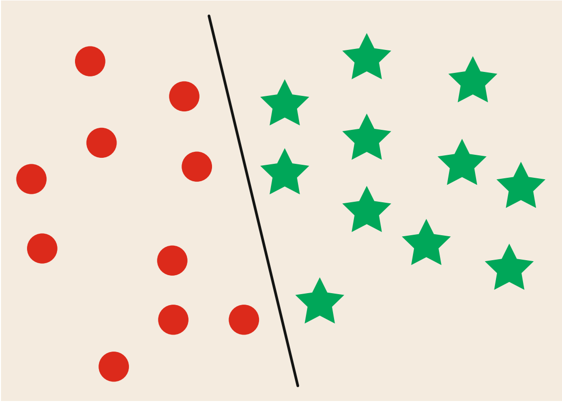


Fig. 36.7 Decision plane which separates the *red circle* class from the *green star* class. Reprinted from [36.10], with permission from Elsevier

outliers to rid our data of unnecessary noise, where the classification is realized with random forests.

Concept-Extraction Using Random Forests (RFs)

Random forests are an ensemble learning method for classification and regression [36.33]. The most prominent characteristic of RFs, which makes them a very popular machine learning algorithm, is their ability to learn multiclass classification problems. RFs operate by constructing a multitude of decision trees at training time. Moreover, RFs operate on two sources of randomness. The first is that each decision tree is grown on a different bootstrap sample drawn randomly from the training data. The second is that at each node split during the construction of a decision tree, a random subset of p variables is selected from the original variable set and the best split based on these p variables is used. When an unknown case arrives, the predictions of all the trees constituting the RF are aggregated using either majority voting for classification or averaging for regression problems. For an RF consisting of N trees, the equation for predicting the class label l of a case y through majority voting is the following

$$l(y) = \operatorname{argmax}_c \left(\sum_{n=1}^{N} I_{h_n}(y) = c \right), \tag{36.9}$$

where I is the indicator function and h_n the n-th tree of the RF.

The RF also provides an internal estimate of its generalization error, by using the out-of-bag (OOB) error estimate. Specifically for each tree that is constructed, 2/3 of the original data instances are used in that particular bootstrap sample, while the remaining instances (OOB data) are classified by the constructed tree and

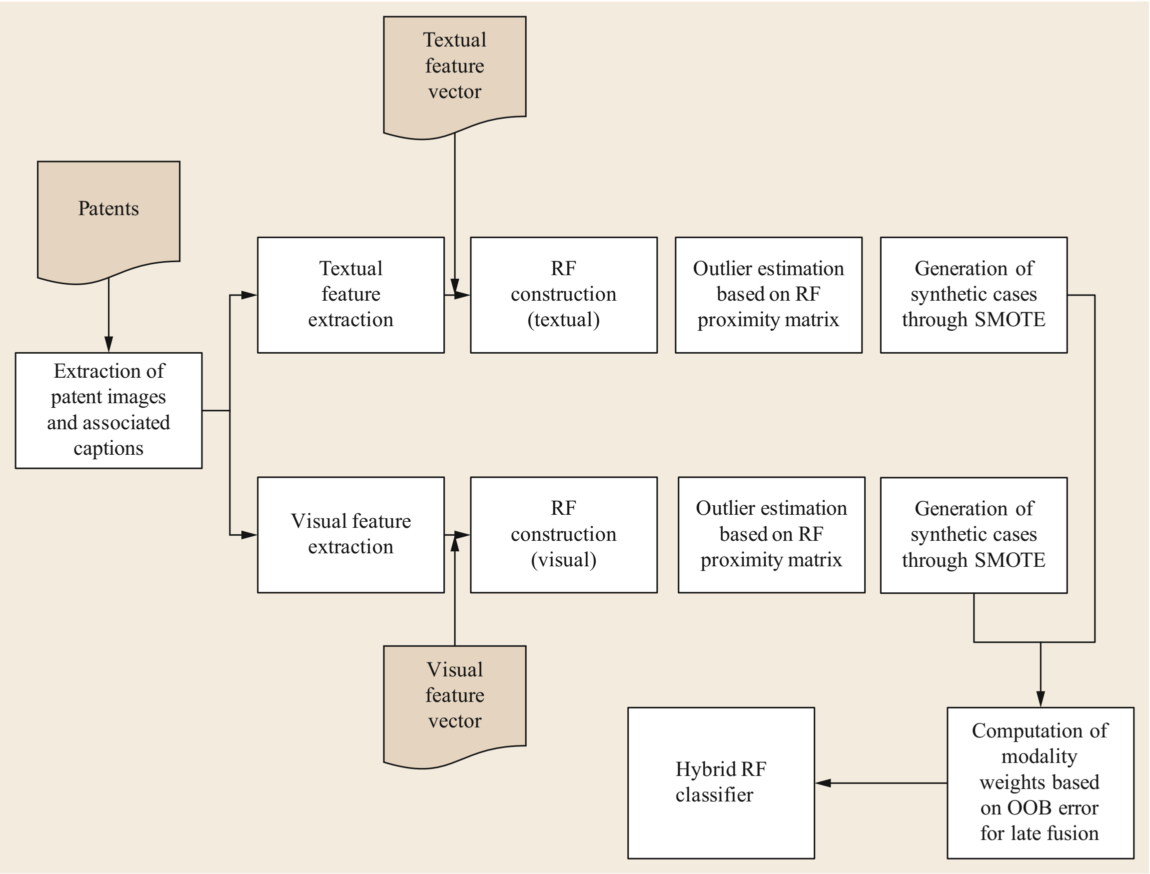


Fig. 36.8 RF multimodal classification for concept extraction in patent images

therefore, used for testing purposes. The OOB error estimate is the averaged prediction error for each training case *y*, using only the trees that do not include *y* in their bootstrap sample. Moreover, the RF has an inherent mechanism for detecting outliers. Within the RF context, instances with very low relevancy to all other instances are considered as outliers [36.33]. Specifically, when an RF is constructed, all the training cases are put down each tree and their proximity matrix is computed, based on whether pairs of cases end up in the same terminal node of a tree. By using this proximity matrix, an outlier measure for each case is derived. The outlier measure of each instance is compared to a specified threshold and if it exceeds it the instance is considered as an outlier. For a more thorough analysis of the core concepts of RFs, see [36.33]. For imbalanced datasets, that is datasets where some of each of the classes are not well represented compared to others, the synthetic minority oversampling technique (SMOTE) is used for constructing efficient classifiers. The basic notion of SMOTE is the synthetic generation of new minority class examples, based on the nearest neighbors of these cases, coupled with the undersampling of the majority class cases. For a more detailed description of the algorithm see [36.34].

The flowchart of the proposed concept extraction and classification framework for the training phase, is depicted in Fig. 36.8. The steps and components of the framework are described as follows.

The first step of the training phase of the framework [36.32] is the extraction of the patent images and the associated caption from the patent documents. In the sequel, the visual and textual features are produced as described in Sect. 36.3. Each modality's features are treated independently, and hence two different feature vectors are formulated. Then the feature vectors from each modality serve as input for the construction of an RF. The detection of outliers is the next step and it is achieved as follows: for each RF, the corresponding dataset's training instances are passed from each tree. If a pair of cases end up in the same terminal node of a tree, their proximity is increased by 1. This is repeated for every pair of cases and all trees in the RF. The final proximity values are calculated by normalizing the computed proximities. Normalization is realized by dividing the proximities values to the number of trees. Thus, if a dataset consists of N cases, an $N \times N$ proximity matrix is derived. From this proximity matrix a measure that indicates the outlierness of each case is computed. Given that the RF algorithm is based on randomization and in order to obtain robust and reliable estimations about potential outliers, it is proposed that the RF construction for the outlier detection and elimination step is repeated several times for each modality and the resulting outlier measure values from the constructed RFs are averaged. The cases that are identified as outliers are eliminated from further processing. The next step targets the problem of imbalanced datasets. As already mentioned the method opted for to solve this issue is the oversampling procedure of SMOTE, which generates in an artificial manner new cases and supplements existing ones. The resulting larger datasets can lead to a better and more efficient RF training. However, we should note that for the particular dataset SMOTE algorithm is solely used for introducing new training instances and not for balancing the dataset given that is balanced by default. According to [36.34], SMOTE oversamples each case by introducing synthetic examples along the line segments joining a number (the number depends on the amount of oversampling required) of that case's nearest neighbors. Oversampling is applied to all datasets and concepts separately. After the application of the SMOTE method, the final datasets are produced and eventually the final RFs for the textual and visual features. The final RF predictions are produced by using a late fusion strategy. Specifically from the OOB error estimate (for the entire data set) of each modality's RF, the corresponding OOB accuracy values are computed. Then, these values are normalized and served as weights for the two modalities.

During the testing phase, the RF provides a probability estimate per class for the new unknown case. This is produced by multiplying the probability outputs P_t and P_v of the textual and visual RFs respectively with their corresponding modality weights W_t and W_v and eventually adding them up, as follows

$$P_{fused} = W_t P_t + W_v P_v \,. \qquad (36.10)$$

36.5.2 Quantitative Evaluation of Concept-Based Patent Search

In order to evaluate the aforementioned concept-based techniques and the classifiers proposed, a dataset was created by extracting figures and their captions from 300 patents selected from the A43B and A63C IPC subclasses. The dataset contains parts of footwear and the data extracted were manually annotated. On the basis of advice from professional patent searchers in that domain, the following eight concepts were selected: cleat, ski boot, high heel, lacing closure, heel with spring, tongue, toe caps, and roller skates. The quantitative evaluation is two fold: firstly, for the demonstration of the superiority of the hybrid approach, compared to the unimodal textual or visual case, and secondly, for the comparison between two classification methods; support vector machines (SVM) and random forests (RFs).

Tables 36.3 and 36.4 contain respectively the results of an SVM-based and an RF-based framework used for classifying patent images. The comparison between the two presented frameworks of Figs. 36.6 and 36.8 is done using precision, recall and F-scores, in Tables 36.3 and 36.4, respectively. Although the frameworks use slightly different fusion strategies, there is some indication of the superiority of RF in the patent image classification domain (in comparison to the SVM classification). At this point, we should note that the fusion strategy followed in the case of the SVM classification is a simple early fusion approach where the descriptor employed for hybrid classification was the result of the concatenation of the textual and visual feature vectors.

Regarding the parameters of the methods involved the following setting is applied for the case of RF classification. The number of trees of the RF was defined

Table 36.3 Precision, recall and F-score for the concept detectors, using the SVM classification [36.10]

Concepts	Visual (%)			Textual (%)			Hybrid (%)		
	Precision	Recall	F-score	Precision	Recall	F-score	Precision	Recall	F-score
Cleat	84.38	45.76	59.34	89.13	69.49	78.10	89.58	72.88	80.37
Ski boot	84.62	67.35	75.00	87.18	69.39	77.27	93.02	81.63	86.96
High heel	82.69	72.88	77.48	76.79	72.88	74.78	92.59	84.75	88.50
Lacing closure	79.17	41.30	54.29	63.64	45.65	53.16	88.46	50.00	63.89
Heel with spring	69.70	54.76	61.33	96.15	59.52	73.53	100.0	45.24	62.30
Tongue	75.68	57.14	65.12	100.0	83.67	91.11	95.12	79.59	86.67
Toe caps	60.53	53.49	56.79	75.68	65.12	70.00	70.21	76.74	73.33
Roller skates	82.50	49.25	61.68	86.15	83.58	84.85	96.55	83.58	89.60
Average	77.41	55.24	63.88	84.34	68.66	75.35	90.69	71.80	78.95

Table 36.4 Precision, recall and F-score for the concept detectors, using the RF classification [36.32]

Concepts	Visual (%)			Textual (%)			Hybrid (%)		
	Precision	Recall	F-score	Precision	Recall	F-score	Precision	Recall	F-score
Cleat	66.1	66.1	66.1	79.2	71.2	75	89.1	83.1	85.9
Ski boot	85.7	73.5	79.1	77.7	85.7	81.5	80.4	83.7	81.9
High heel	68.6	81.4	74.4	76.9	67.8	72	80.6	84.7	82.6
Lacing closure	50.0	76.1	60.3	42.4	60.9	50.0	67.3	76.1	71.4
Heel with spring	68.1	71.4	69.7	73.9	81	77.3	90.2	88.1	89.1
Tongue	78.3	59.2	67.4	86.3	89.8	88	88.2	91.8	89.9
Toe caps	72.2	60.5	65.8	90.6	67.4	77.2	89.7	81.4	85.3
Roller skates	74.2	73.1	73.6	90	80.6	85	90.5	85.1	87.7
Average	70.4	70.1	69.5	77.1	75.5	75.7	84.5	84.2	84.2

by the OOB error estimate. Thus, after conducting several experiments which involved gradual increase of the number of trees used, it was found that the OOB error estimate was fairly stable after when the number of trees was 1000. Adding more trees did not translate to further improvement of OOB. Hence, the number of trees was set to 1000. Moreover, the number of the subset of variables used to determine the best split during the growing of each tree was set to $\sqrt{k}$, where k is the total number of features of the dataset [36.33]. Regarding the RF outlier detection according to [36.35], a case can be considered an outlier if its outlier measure value is higher than 10. The result of this configuration was that around 2% of the textual modality's cases were detected as outliers and discarded, while for the visual modality no outliers were detected. Finally, the SMOTE oversampling rate for each concept in both modalities datasets was set to 500%, which is equivalent to the generation of 5 new synthetic cases for each real case.

36.5.3 Qualitative Evaluation Through a Patent Search System

One of the first systems that dealt with patent image search was PATSEEK [36.23], while later the PatMedia [36.9] patent image search engine was developed. We illustrate visual results of concept detection through PatMedia in Fig. 36.9. When both features (i. e., visual and textual) are combined, the results are improved as 100% precision is achieved in the top 18 retrieved images. This means that the three erroneous results that appeared in searching only by text or by image, have been assigned a lower score and ranked more than the last retrieved image of the PatMedia interface.

As an illustrative example, the PatMedia [36.9] graphical user interface (GUI) is demonstrated in Fig. 36.10, where searching by concept (Sect. 36.5) or by patent image (Sect. 36.4) are some of the involved functionalities. The quantitative comparison is presented in [36.10], and the PatMedia framework is reviewed in [36.11], while this chapter aims to summarize the overall concept-based patent search, fusing both visual and textual information, supervised machine learning techniques and local binary features.

The user may upload an image to retrieve similar images, in a CBIR scenario, when searching by content. Searching is also available in the case of a textual keyword or in the hybrid scenario. The PatMedia tool integrates all methods that are reviewed in this chapter and serves also as a standalone patent search engine.

Apart from concept extraction techniques in patent images, that are used to retrieve patents through searching by concept, it is also popular to search by pattern and or shape, where effective binary visual descriptors are required. In this regard, we present in detail, in Sect. 36.4, the visual descriptors of Sect. 36.3, known as AHDH, which allow searching by content in an unsupervised manner. The AHDH binary features were introduced in [36.8] as visual descriptors for binary images, and they have been evaluated in the context of patent image retrieval.

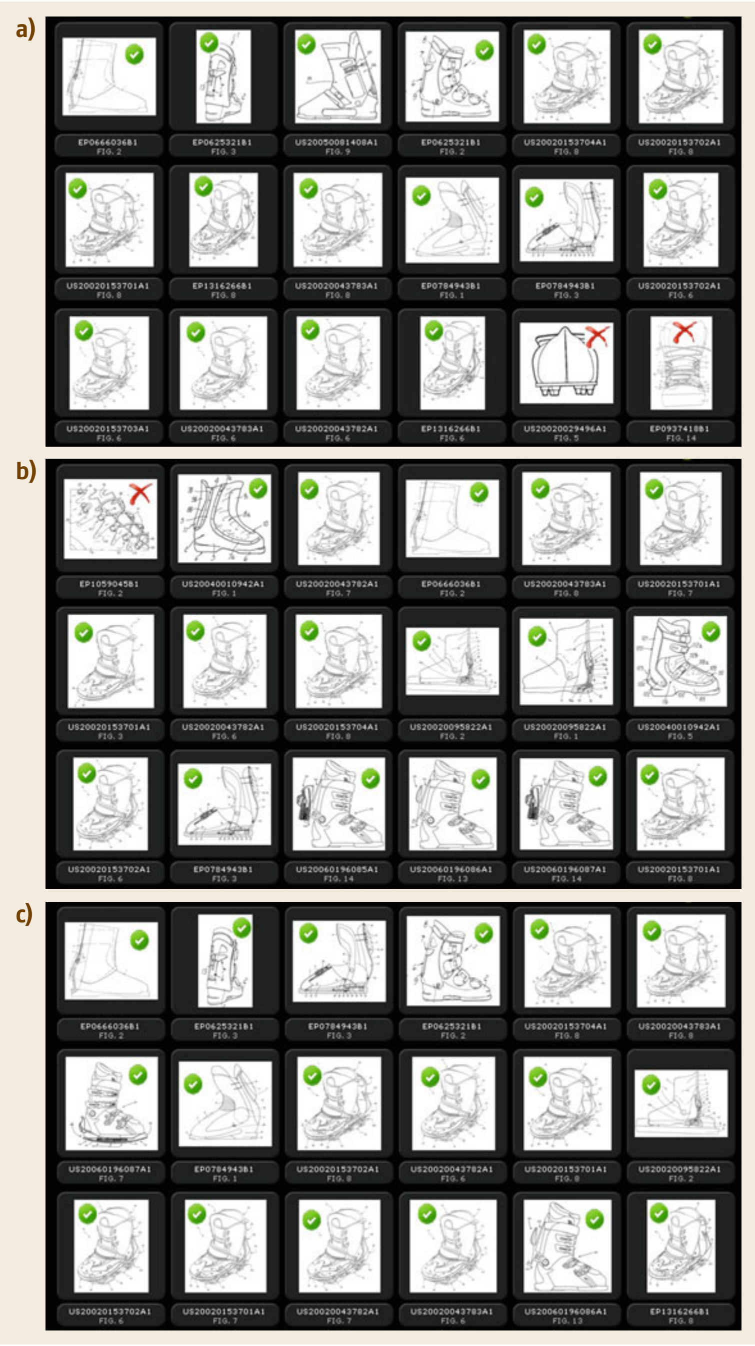


Fig. 36.9a–c PatMedia retrieved results. **(a)** Results for *ski boots* using textual features; **(b)** results for *ski boots* using visual features; **(c)** results for *ski boots* using the hybrid approach. The *green tics* indicate that the results retrieved are correct while the *red X* indicates the wrong results

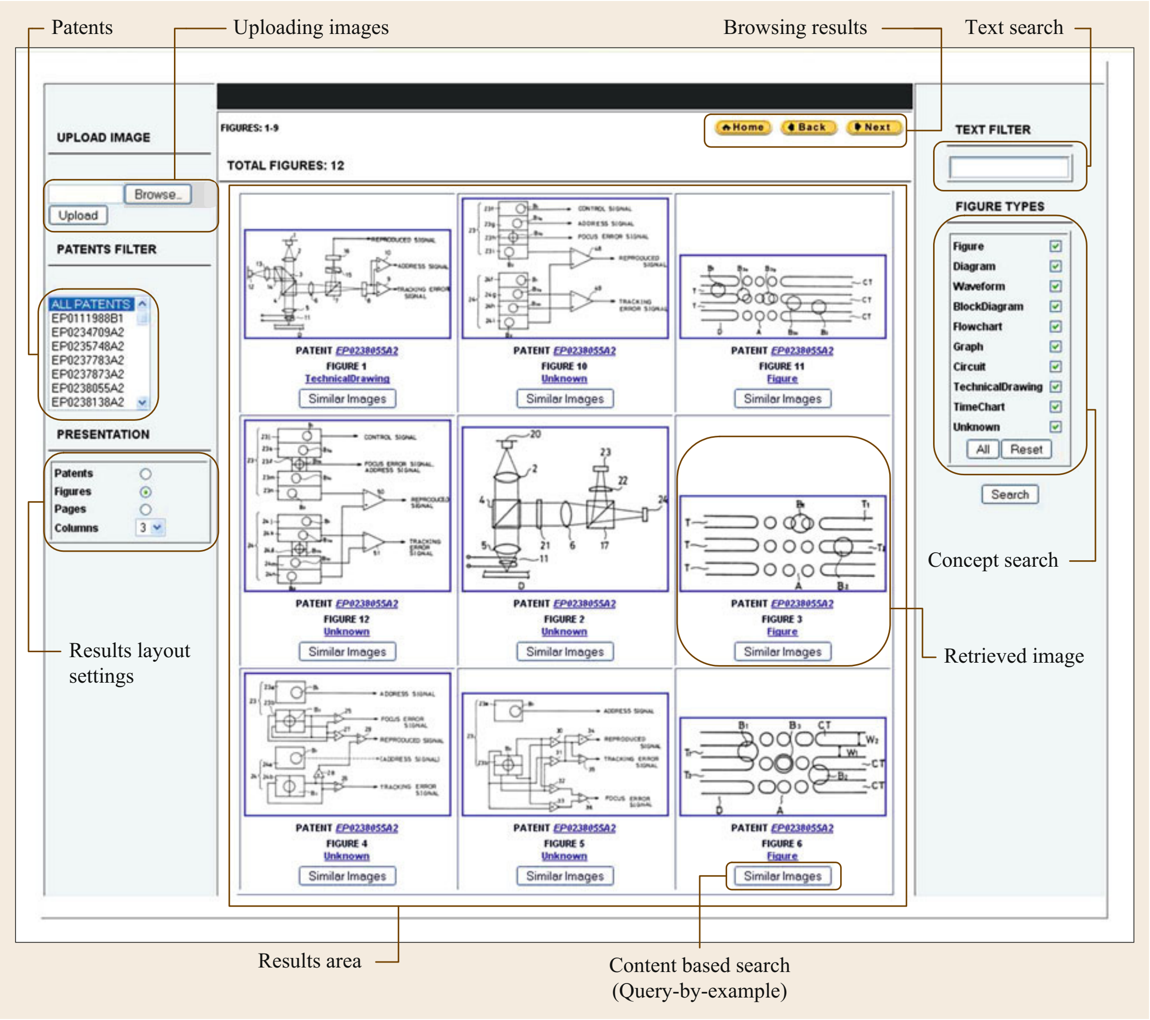


Fig. 36.10 PatMedia graphical user interface (GUI)

36.6 Conclusion

The growing number of patent applications submitted worldwide necessitates the development of advanced patent search technologies. Effective and fast patent search relies on content-based and concept-based patent image retrieval while also fusing both textual and visual modalities that appear in each patent. By considering the visual content of patent images, we facilitate and improve the performance of patent search tasks such as patent invalidation and competitive intelligence research. In this chapter, we described state-of-the-art techniques in both directions, as well as their demonstration within user-friendly patent search systems and tools. Both techniques could be incorporated into existing official patent search systems that are mainly text-based and help them in overcoming the limitations imposed by using strictly textual descriptions which may be incomplete or in different languages.

In regard to content-based techniques, an algorithm called AHDH which is oriented towards handling patent images is described. The AHDH algorithm produces a visual feature representation of the patent images which considers their special characteristics and achieves very good performance compared to other similar algorithms. However, the major limitation of the algorithm, which seems to be the main factor restraining its performance, is the absence of inherent geometrical invariance. Therefore, further research on the AHDH algorithm should move in the geometrical invariant direction. Moreover, given that centroid hierarchical partition produces a plethora of shape-based statistical characteristics, such as density histograms, centroid locations, and blank regions, another idea with potentially interesting results would be to combine these vectors through a machine learning algorithm in order to further improve the algorithms performance.

Following the trend of modern image retrieval towards concept-based image search, the AHDH algorithm is used in a concept-based patent search framework. Different machine learning algorithms were evaluated, including SVM and RF and the experiments realized indicated that RF outperform SVM. Although the framework was for a limited set of concepts, the methodology presented using RF is scalable and the application of SMOTE minimizes the need for training data. Moreover, given that we can limit the search within patents of the same IPC class, the framework can target each time only a specific set of concepts relevant to the corresponding IPC class. A drawback of concept extraction techniques is that they require a training set for each new concept introduced, hence there is a need to have manually annotated images by experts.

Finally, from an application point of view both the concept retrieval and the content-based retrieval using query-by-example method modules could be parts of a larger patent retrieval framework. Such frameworks can also include other functionalities including full text and semantic search for the whole patent document. An example of such a framework is the PatMedia search engine.

References

36.1 D.G. Lowe: Distinctive image features from scale-invariant keypoints, Int. J. Comp. Vis. **60**(2), 91–110 (2004)

36.2 H. Bay, A. Ess, T. Tuytelaars, L. Van Gool: Speeded-up robust features (SURF), Comput. Vis. Image Underst. **110**(3), 346–359 (2008)

36.3 M. Rusinol, L.-P. de las Heras, O.R. Terrades: Flowchart recognition for non-textual information retrieval in patent search, Inf. Retr. **17**(5/6), 545–562 (2014)

36.4 K. Xu, H. Lin, Y. Lin, B. Xu, L. Yang, S. Zhang: Patent retrieval based on multiple information resources, Inf. Retr. Technol. **9994**, 125–137 (2016)

36.5 W. Tannebaum, A. Rauber: Learning keyword phrases from query logs of USPTO patent examiners for automatic query scope limitation in patent searching, World Patent Inf. **41**, 15–22 (2015)

36.6 M. Mogharrebi, M.C. Ang, A.S. Prabuwono, A. Aghamohammadi, K.W. Ng: Retrieval system for patent images, Procedia Technol. **11**, 912–918 (2013)

36.7 F. Piroi, A. Hanbury: Evaluating information retrieval systems on European patent data: The CLEF-IP campaign, Curr. Chall. Patent Inf. Retr. **37**, 113–142 (2017)

36.8 P. Sidiropoulos, S. Vrochidis, I. Kompatsiaris: Content-based binary image retrieval using the adaptive hierarchical density histogram, Pattern Recognit. **44**(4), 739–750 (2011)

36.9 S. Vrochidis, S. Papadopoulos, A. Moumtzidou, P. Sidiropoulos, E. Pianta, I. Kompatsiaris: Towards content-based patent image retrieval: A framework perspective, World Patent Inf. **32**(2), 94–106 (2010)

36.10 S. Vrochidis, A. Moumtzidou, I. Kompatsiaris: Concept-based patent image retrieval, World Patent Inf. **34**(4), 292–303 (2012)

36.11 S. Vrochidis, A. Moumtzidou, I. Kompatsiaris: Enhancing patent search with content-based image retrieval, Prof. Search Mod. World **8830**, 250–273 (2014)

36.12 S. Vrochidis, A. Moumtzidou, G. Ypma, I. Kompatsiaris: PatMedia: Augmenting patent search with content-based image retrieval. In: *Inf. Retr. Facil. Conf.* (2012) pp. 109–112

36.13 D. De Marco, A. Davis: *Mechanical Patent Searching: A Moving Target* (Patent Information Users Group (PIUG), Baltimore 2010)

36.14 J. Sas, U. Markowska-Kaczmar, A. Moumtzidou: Caption-guided patent image segmentation. In: *Fed. Conf. Comp. Sci. Inf. Syst. (FedCSIS) 2016* (2016) pp. 261–270

36.15 M.F. Porter: An algorithm for suffix stripping, Program **14**(3), 130–137 (1980)

36.16 A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain: Content-based image retrieval at the end of the early years, IEEE Trans. Pattern Anal. Mach. Intell. **22**(12), 1349–1380 (2000)

36.17 F. Long, H. Zhang, D.D. Feng: Fundamentals of content-based image retrieval, Multimed. Inf. Retr. Manag. 1–26 (2003)

36.18 R. Datta, D. Joshi, J. Li, J.Z. Wang: Image retrieval: Ideas, influences, and trends of the new age, ACM Comp. Surv. (Csur) **40**(2), 5 (2008)

36.19 P. Dukkipati, L. Brown: Improving the recognition of geometrical shapes in road signs by augmenting the database. In: *Proc. 3rd Int. Conf. Comp. Sci. Appl.* (2005) pp. 8–13

36.20 I. Yahiaoui, N. Hervé, N. Boujemaa: Shape-based image retrieval in botanical collections. In: *Pacific-Rim Conf. Multimed.* (2006) pp. 357–364

36.21 S. Antani, L.R. Long, G.R. Thoma: Content-based image retrieval for large biomedical image archives. In: *Proc. 11th World Congr. Med. Inform. (MEDINFO)* (2004) pp. 7–11

36.22 N. Alajlan, I.E. Rube, M.S. Kamel, G. Freeman: Shape retrieval using triangle-area representation and dynamic space warping, Pattern Recognit. **40**(7), 1911–1920 (2007)

36.23 A. Tiwari, V. Bansal: PATSEEK: Content based image retrieval system for patent database. In: *ICEB* (2004) pp. 1167–1171

36.24 M. Yang, G. Qiu, J. Huang, D. Elliman: Near-duplicate image recognition and content-based image retrieval using adaptive hierarchical geometric centroids. In: *18th Int. Conf. Pattern Recognit., ICPR 2006*, Vol. 2 (2006) pp. 958–961

36.25 F. Mahmoudi, J. Shanbehzadeh, A.-M. Eftekhari-Moghadam, H. Soltanian-Zadeh: Image retrieval based on shape similarity by edge orientation autocorrelogram, Pattern Recognit. **36**(8), 1725–1736 (2003)

36.26 G. Zhu, X. Yu, Y. Li, D. Doermann: Learning visual shape lexicon for document image content recognition, Comp. Vis.–ECCV **5303**, 745–758 (2008)

36.27 C. Bauckhage, J.K. Tsotsos: Bounding box splitting for robust shape classification. In: *IEEE Int. Conf. Image Process., ICIP 2005*, Vol. 2 (2005) p. II–478

36.28 J.P. Eakins, K.J. Riley, J.D. Edwards: Shape feature matching for trademark image retrieval. In: *Int. Conf. Image Video Retr.* (2003) pp. 28–38

36.29 C. Bauckhage: Tree-based signatures for shape classification. In: *IEEE Int. Conf. Image Process.* (2016) pp. 2105–2108

36.30 R. Mörzinger, A. Horti, G. Thallinger, N. Bhatti, A. Hanbury: Classifying patent images. In: *CEUR Workshop Proc.*, Vol. 1177 (2011)

36.31 A. Moumtzidou, S. Vrochidis, I. Kompatsiaris: Concept extraction from patent images based on recursive hybrid classification. In: *Inf. Retr. Facil. Conf.* (2013) pp. 83–86

36.32 D. Liparas, A. Moumtzidou, S. Vrochidis, I. Kompatsiaris: Concept-oriented labelling of patent images based on random forests and proximity-driven generation of synthetic data. In: *V&L Net*, Vol. 25 (2014)

36.33 L. Breiman: Random forests, Mach. Learn. **45**(1), 5–32 (2001)

36.34 N.V. Chawla, A. Lazarevic, L.O. Hall, K.W. Bowyer: SMOTEBoost: Improving prediction of the minority class in boosting. In: *Eur. Conf. Princ. Data Min. Knowl. Discov* (2003) pp. 107–119

36.35 L. Breiman: Manual–setting up, using, and understanding random forests V 4.0, https://www.stat.berkeley.edu/~breiman/Using_random_forests_v4.0.pdf (2003)

Part E | 36

Ilias Gialampoukidis

Information Technologies Institute
Centre for Research and Technology Hellas
Thermi, Greece
heliasgj@iti.gr

Ilias Gialampoukidis is a Postdoctoral researcher at the Centre for Research and Technology Hellas – Information Technologies Institute. He received a PhD in Mathematics in 2014 and an MSc in Statistics and Modelling in 2011 from the Aristotle University of Thessaloniki, Greece. His research interests involve information retrieval from textual and visual content, big data analytics, multimodal fusion, supervised and unsupervised learning, and social media network analytics.

Anastasia Moumtzidou

Information Technologies Institute
Centre for Research and Technology Hellas
Thermi, Greece
moumtzid@iti.gr

Anastasia Moumtzidou has been a Research Assistant at the Information Technologies Institute (ITI) of the Centre for Research and Technology Hellas (CERTH) since January 2007. Her research interests include web application technologies, database system management, information retrieval, image and video search, search engine development, as well earth observation data analysis. She has co-authored more than 60 journal / conference publications and book chapters.

Stefanos Vrochidis

Information Technologies Institute
Centre for Research and Technology Hellas
Thermi, Greece
stefanos@iti.gr

Stefanos Vrochidis is a Senior Researcher at the Information Technologies Institute of the Centre for Research and Technology Hellas. His research interests include multimedia analysis, web and social media mining, semantics, information retrieval, image and video search, multimodal analytics, decision support, as well as security and robotics applications. He has co-authored more than 110 journal and conference publications.

Ioannis Kompatsiaris

Information Technologies Institute
Centre for Research and Technology Hellas
Thermi, Greece
ikom@iti.gr

Ioannis Kompatsiaris is Research Director at CERTH-ITI, the Head of the Multimedia Knowledge and Social Media Analytics Laboratory and Deputy Director of the Institute. His research interests include multimedia, big data and social media analytics, semantics, human computer interfaces (AR and BCI), eHealth, security and culture applications.He is the co-author of 129 papers in refereed journals, 46 book chapters, 8 patents and more than 420 papers in international conferences.

37. Methodological Challenges for Creating Accurate Patent Indicators

Ulrich Schmoch, Mosahid Khan

The chapter deals with new methodological issues of retrieval for patent indicators linked to the change of the patent system in the last 20 years and the new ways to access patent data. In particular, it describes international flows of patent applications between the US, Europe, and Southeast Asia, and illustrates methods for an appropriate cross-country comparison. A central topic of this chapter is the implications of the frequently used Patent Cooperation Treaty (PCT) route of patent applications on the conception of search strategies and the interpretation of search results. Furthermore, the possibilities of search with the new international Cooperative Patent Classification (CPC) are explained. In addition, the patenting activities of very large companies and patent value are discussed.

37.1 New Methodological Issues

It used to be quite difficult and expensive to get access to patent data. In the recent past, various patent databases have been provided by private vendors or public authorities at low cost or that are even free and with comfortable conditions of technical use. Examples are the European Patent Office's (EPO) statistical database PATSTAT and their search platform Espacenet, as well as the patent search tool Patentscope of the World Intellectual Patent Office (WIPO). This situation tempts many researchers to conduct quick and suboptimal patent searches, and they are not aware of the inanity of their results. Many basic aspects of the methodology of patent indicators were already alluded to in the former version of this handbook, in *Hinze* and *Schmoch* [37.1]. Since then, new methodological issues have emerged. Against this background, we will address the most important new stumbling blocks in patent searches and in creating accurate patent indicators. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Fraunhofer Institute for Systems and Innovation Research or the World Intellectual Property Organization.

37.2 International Patent Flows

As a basic principle, granted patents are territorial rights that only apply to the country for which the patents have been granted. For example, if a patent is granted in China, it will not be enforceable in Japan unless a patent for the same invention is applied for and granted in Japan. The legal system and the economic and geopolitical situation differ substantially by country. For instance, the scope of protection of an application in Japan is smaller than in Germany, implying a much higher number of patent applications

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_37

Table 37.1 Patent applications by residents and industrial R&D expenditures for selected countries, 2013 (after [37.2, 4, 5])

	Patent applications by residents	Industrial R&D (2010 million dollars in constant prices and PPPs)	Ind. R&D/appl.
Germany (DPMA)	50 414	64 259	1.27
US (USPTO)	287 831	305 311	1.06
Japan (JPO)	271 731	117 571	0.43
South Korea (KIPO)	159 978	53 507	0.33
China (SIPO)	704 936	214 324	0.30

per expenditure on research and development in Japan compared to Germany; South Korea has a strong orientation of external patenting towards the United States, whereas European countries have a primary external orientation towards other European countries, etc. In addition, the propensity to patent varies across country, industry, and firm size. Therefore, comparisons of patent application data across offices should be undertaken with care. However, many providers offer easy access to databases with a broad coverage of national patent applications from about 50 patent authorities. The databases are generally derived from the file INPADOC, formerly offered by a private company, at present by the EPO. The most popular version of such an encompassing file is PATSTAT, offered by the EPO at low cost. Therefore, many users consider all available data without paying attention to the differences of countries and industries.

Furthermore, despite harmonization of patent law (e. g., the Trade-Related Aspects of Intellectual Property Rights (TRIPS) agreement), patent law still differs across offices. A search for a specific topic in patent databases without defining basic criteria will lead to an inscrutable mix of applications from different patent systems. In particular, the average economic value of a patent application at different offices can vary considerably. In addition, the value of patents varies within an office. The application behaviors at national offices often differ considerably, even if the formal patent laws seem to be quite similar. Different application fees, different incentives of the government or of the employers, different interpretations of claims, or different patent cultures can be the underlying reasons. The considerable implications for statistics can be seen in a comparison of domestic patent applications at selected offices (Table 37.1).

The technological innovation activity of a country is closely linked to its research and development (R&D) by industry. Therefore, the ratio of industrial R&D to the number of patent applications of residents may be considered as an indicator for the average value of a patent application. The comparison of the US, Germany, Japan, South Korea, and China shows a higher average level for Germany and the US and a lower one for Japan, South Korea, and China. Regarding patent statistics, this implies that counting all patent applications worldwide without any specification generates meaningless data, as domestic systems with quite a high

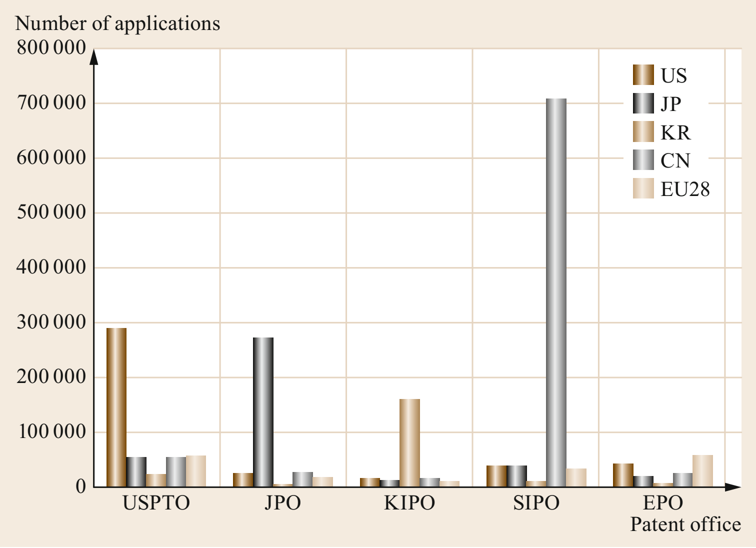


Fig. 37.1 Number of applications at IP5 offices by applicants from IP5 countries, 2013 (after [37.2, 3])

Part E | 37.2

propensity to patent dominate the results without a clear link to economic or technological value.

The consequences of this application behavior may be demonstrated by the number of applications by different countries at the so-called IP5 offices. IP5 is a forum of the five largest intellectual property offices, namely (Fig. 37.1):

1. China (State Intellectual Property Office of the People's Republic of China—SIPO)
2. Japan (Japan Patent Office—JPO)
3. South Korea (Korean Intellectual Property Office—KIPO)
4. The US (United States Patent and Trademark Office—USPTO)
5. The European Patent Office (EPO).

From Fig. 37.1, it is obvious that the domestic countries substantially dominate the application numbers at their domestic office, e. g., US applicants at the USPTO or Chinese applicants at SIPO. From a statistical perspective, this is called domestic advantage. At the EPO, a less important regional advantage of countries in the European Union (EU) becomes visible as well. Furthermore, there exist regional preferences of the countries of origin. For instance, Japanese or Korean applicants file more applications at the USPTO than at the SIPO or the EPO, or EU applicants prefer the SIPO over the KIPO. Thus, without any confinement, the search results will be dominated by domestic filings. For instance, 85% of Chinese applications at the IP5 offices are applied for at the SIPO; in contrast only 70% of the US applications are applied for at the USPTO. In consequence, the distortion between countries illustrated in Table 37.1 is directly visible in patent statistics.

The search for appropriate systems for country comparisons has always been a major concern of patent methodology. Various concepts were presented in *Hinze* and *Schmoch* [37.1]. The most relevant approach in the 1990s and at the beginning of the 2000s was the triad concept focusing on Japan, the US, and Europe [37.1, 6]. However, with the substantial growth in patent filings originating from Southeast Asia, in particular China and South Korea, the results of the triad approach are no longer considered appropriate. This change is directly reflected in the patent flows between the major countries, represented by the so-called IP5 offices.

Examining the present patent flows between IP5 offices, applications to the EPO originating from the US are higher than those from Japan and China and particularly those from South Korea (Fig. 37.2). However, the number of applications from these three Southeast Asian countries combined (CJK) prove to be clearly

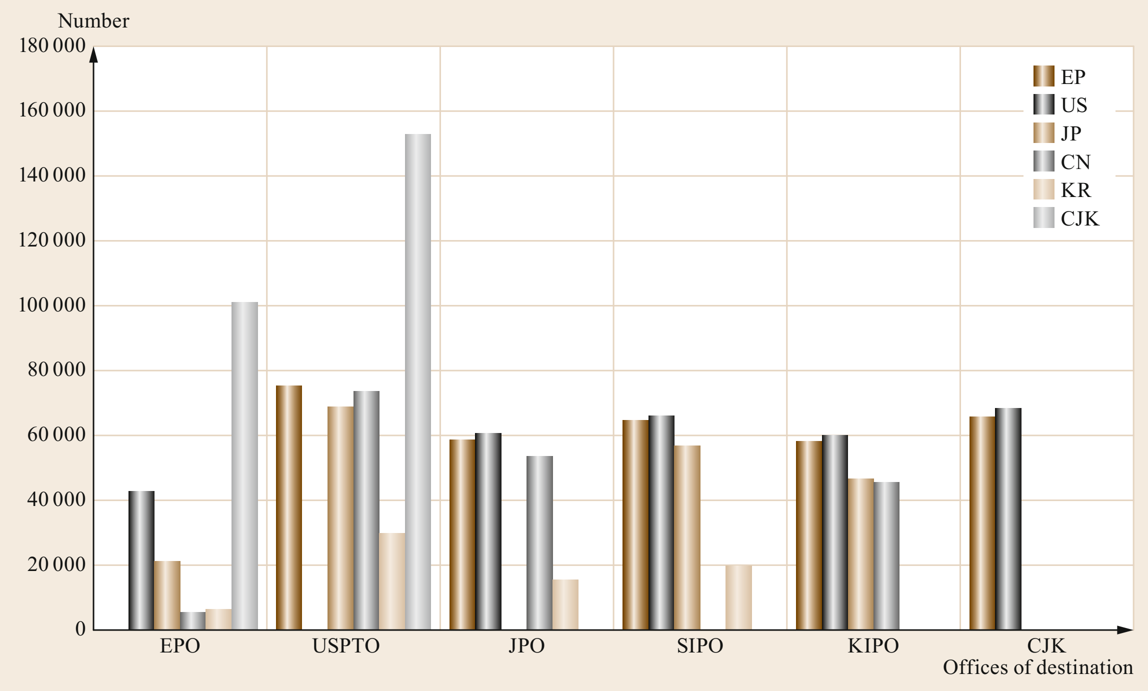


Fig. 37.2 Number of total patent applications from selected countries/regions of origin to selected external patent offices, 2014 (after [37.5])

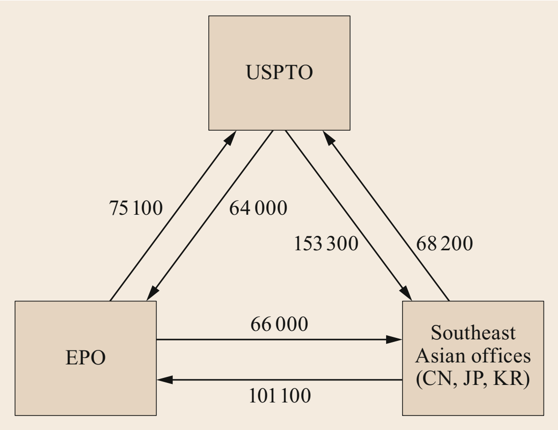


Fig. 37.3 Patent flows between offices of the new triad, 2014 (after [37.5])

higher than those from the US. The Southeast Asian countries exhibit a special interest in the US market. The flows from Europe and the US to the three Southeast Asian offices (CJK) prove to be nearly equivalent. Therefore, the combined flow from Europe (EU countries) or the US to at least one of the three Southeast Asian offices is only slightly higher than the flow to one of the offices.

Against this background, it is easy to conceive a new triad model with the three poles USPTO, EPO and large Southeast Asia Offices (SIPO, JPO, KIPO) combined. The resulting flows are illustrated in Fig. 37.3. In this graph, the substantial flows from Southeast Asia to Europe and particularly to the US become obvious.

Looking at the development over time, Fig. 37.4 documents the enormous increase of the flows from Southeast Asia to Europe and the US between 2000 and 2014, primarily linked to the growth of flows from China and—to a lower extent—from South Korea, and more recently from Japan. The flows from Europe (EP) and the US to Southeast Asia (CJK, JPO, SIPO, and KIPO) have increased as well, but to a lower extent.

This new international structure requires new approaches to achieve appropriate data for country comparisons replacing the outdated old triad concept. The methodological basis of such an approach must be to achieve samples with comparable economic values of the patents compared. For this purpose, it is useful to consider the costs of patent applications in more detail.

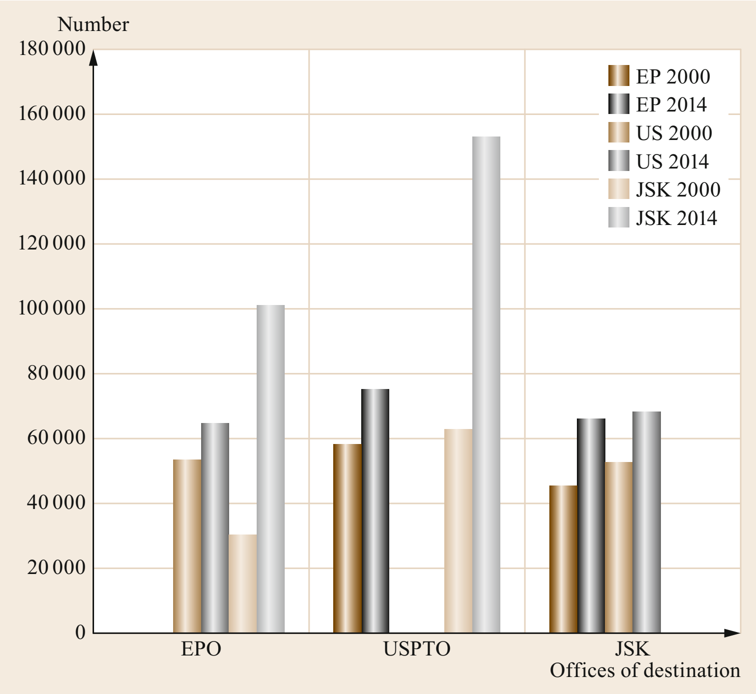


Fig. 37.4 Patent flows between offices of the new triad, 2000 and 2014 (after [37.5])

37.3 Costs of Patent Applications

The costs of patent applications depend on a variety of different factors, and it is important to consider all elements in order to become aware of the relevant ones. Table 37.2 lists the most important costs linked to a patent application.

The official fees in a country comparison were analyzed in detail by *Van Pottelsberghe* and *Danguy* [37.7] and *De Rassenfosse* and *Van Pottelsberghe* [37.8].

To give an impression of the order of the magnitude, the official fees for application, examination, and renewal are shown in Table 37.3 for the IP5 offices.

Table 37.2 Major cost elements of patent applications

Official fees
Application fees
Search fees, examination fees
Grant fees
Annual renewal fees
Foreign application fees
Foreign examination fees
Procedural fees
Patent attorneys' costs
Draft of patent application
Studying the search results and the examiner's reports, answering the examiner's report
Monitoring of the regular payment of renewal fees
Organizing the transfer to national offices in the case of an EU application (identifying correspondence attorneys, translating the applications)
Organizing applications in foreign countries and/or PCT applications at the end of the priority year and translating the original applications
Attending and acting in the examination process of the applications in foreign countries
Studying the relevant documents in case of opposition, answering the examiner's questions

The official fees can be assessed as generally moderate. The search and examination fees at the EPO are distinctly higher than at the USPTO or Deutsches Patent- und Markenamt (DPMA, German Patent and Trademark Office). In all five cases, the maintenance fees increase progressively and achieve high values especially at the USPTO (at the USPTO, the maintenance fees do not have to be paid annually but every 4 years (according to the USPTO, patents are renewed at 3.5, 7.5, 11.5 years, etc.). To make the data comparable with other offices, the average per year was calculated in Table 37.3). Consequently, only very valuable patents will be maintained up to the maximum limit of 20 years [37.9].

The costs of the patent attorney primarily concern the draft of the application and the legal assistance in the search and examination procedure. These costs largely depend on the complexity of the application, ranging from extremely simple to highly complex, e. g., from an electric switch to an MRI (magentic resonance imaging) scanner. Estimates for patent attorneys' costs are given in Table 37.4 (the cost estimates are based on the specifications of the websites of various patent attorneys).

In the total costs of patenting, the shares of the patent attorneys' costs are generally larger than those of the official fees. Due to the plurality of cost items, only rough estimates are possible as to the total costs of a patent application: The costs of an application with a period of validity of 10 years will be in the range of €20 000–30 000 for a national application, a European application with 5 designated countries and a realised transfer to these countries may achieve a level of €70 000–90 000, an international application with transfer to the EPO and 5 designated European countries and transfer to a further 5 non-European countries

Table 37.3 Selected official fees at five selected offices (as of September 2016)

Currency	USPTO $	EPO € ($)[a]	JPO Yen ($)	SIPO Yuan Renminbi (CNY) ($)	KIPO KRW ($)
Application fees (online)	280	120 (133)	14 000 (129)	950 (143)	66 000 (57)
Search fees	600	1300 (1468)	–	–	–
Examination fees	720	1635 (1809)	118 000 (1085)	2500 (376)	p.cl. 143 000 + 44 000 (125 + 38 p.cl.)
Maintenance fees, year 4	450	585 (647)	6400 (59)	1200 (181)	p.cl. 40 000 + 20 000 (34 + 17 p.cl.)
Maintenance fees, year 12	1850	1575 (1742)	55 400 (509)	4000 (602)	p.cl. 240 000 + 55 000 (207 + 47 p.cl.)

[a] Conversion rates—purchasing power parities (ppp)—OECD data
Source: Websites of the offices

Table 37.4 Estimates of patent attorneys' costs for selected items (as of September 2016)

Procedural step	USPTO $	DPMA € ($)	EPO € ($)
Draft of patent application	5000–16 000	2000–5000 (2200–5500)	3000–6000 (3300–6600)
Search and examination procedure	1000–3000	1000–5000 (1100–5500)	1000–5000 (1100–5500)

Source: Websites of various US American and European patent attorneys, selection of most frequently indicated costs

will reach total costs of about €120 000. The relatively high costs of European and international applications compared to national ones are due to the fact that these applications are finally transferred to several national offices.

In conclusion, it is obvious that the costs of achieving and maintaining a patent are considerable and are not spent just for the honour of having a patent. Without the expectation of substantial economic returns, a patent application will not be filed.

37.4 Patent Applications to Foreign Countries

A further important element for understanding the structure of the international patent system is the legal basis and typical paths of applications to foreign countries.

For statistical analyses, the first application—the so-called priority application—is the most important one, as it is the point in time nearest to the time where the invention was generated. In some countries, it is possible to submit modifications of the application, which adds new embodiments of the priority application, e. g., the continuation-in-part applications at the USPTO or the divisional applications at the EPO. Consequently, an application can have two or more priority dates, the original and the subsequent ones. To avoid double counting of the same invention in a statistical analysis, only the first priority date should be counted.

The date of priority is the relevant point in time for defining novelty. An identical or very similar publication of the object of application earlier than the date of priority implies a rejection of the application.

An important action after the filing of the priority application is the application for the same invention in several countries within the so-called priority year. In most cases, the priority application is submitted to the domestic patent office. However, there are some exceptions. For example, in the first instance, Canadian applicants tend to file at the USPTO, while Swiss applicants prefer to file at the EPO. In principle, it is necessary to file a patent application to every country where protection is striven for. However, some exceptions exist, where the multi-country application process is simplified:

- Patent applications filed at the EPO
- The international patent applications filed via the PCT
- Patent applications filed at the OAPI (Organisation Africaine de la Propriété Intellectuelle)
- Patent applications at the ARIPO (African Regional Intellectual Property Organization)
- Patent applications filed at the EAPO (Eurasian Patent Organization)
- Patent applications filed at the GCCPO (Gulf Cooperation Council Patent Office).

To benefit from the priority claim of first application, applicants need to submit applications at foreign jurisdiction within 12 months after the priority date.

For the present context, patent applications filed at the EPO and the international patent applications filed via the PCT treaty are relevant and discussed in more detail.

The EPO is a regional patent office that was established in 1978 under the European Patent Convention (EPC). It is a centralized system for applications and granting of European patents for the EPC member states. Currently, there are 38 members—28 EU members plus 10 others, including Switzerland and Turkey. The EPO is responsible for granting European patents, but patents granted by the EPO must be validated at national patent offices to be enforceable in those jurisdictions. National offices may require translation of European patents into one of its official languages and additional fees for publication. Applicants decide whether to validate EPO patents in all EPC member states or a selection of member states. On average, EPO patents are validated in the three large economies, namely France, Germany, and the UK. To summarize, at the EPO, the application and the examination is centralized, the final protection refers to special selected countries and is national.

The application at the OAPI refers to 16 member countries of the Francophone part of Africa. The office can grant regional patents on behalf of the member states, thus the OAPI is a central application and examination office. The ARIPO represents 19 African countries, most of them anglophone. It is a central authority for receiving applications. The EAPO centrally receives and grants Eurasian patents valid on the territory of the ten member-states of the Eurasian patent convention, all of them former member countries of the Soviet Union. The GCCPO receives and grants patents valid in the GCC (Gulf Cooperation Council) member countries.

The PCT is an international patent law treaty that was concluded in 1970. The PCT system facilitates the filing of patent applications worldwide and makes it possible to seek patent protection for an invention simultaneously in each of a large number of countries by first filing a single international patent application. The granting of patents remains under the control of national or regional patent offices. The application is made at a receiving office (RO) in one language. The next step is a search performed by an International Search Authority (ISA), resulting in a search report with a written opinion regarding the patentability of the invention that is the subject of the application. It is optionally followed by a preliminary examination, performed by an International Preliminary Examining Authority (IPEA). The ROs, ISAs, and IPEAs are selected at existing patent offices, e.g., the USPTO or the EPO, working on behalf of the WIPO. In the final step, the inventor/applicant must decide whether he/she will pursue the application and in which countries he/she will really seek patent protection. A possible option within PCT is also an application at the EPO, the so-called Euro-PCT application. The international application has a variety of advantages. In particular:

- The applicant can submit valid applications designated for foreign countries shortly before the end of the priority year, as a translation into a foreign language, and the nomination of correspondence patent attorneys in the designated states is no longer necessary, while it is required in the standard procedure without PCT.
- The applicant receives additional information on the patentability of the application through the reports of the ISAs and IPEAs before making the decision on cost-intensive applications in foreign countries.
- The international phase ends 30 months from the priority date and, therefore, the applicant has to make the decision concerning applications in foreign countries much later than the end of the priority year. At 2.5 years after the priority application, more information is available on whether the invention will be successful in the market and will justify the costs of patent applications in foreign countries.

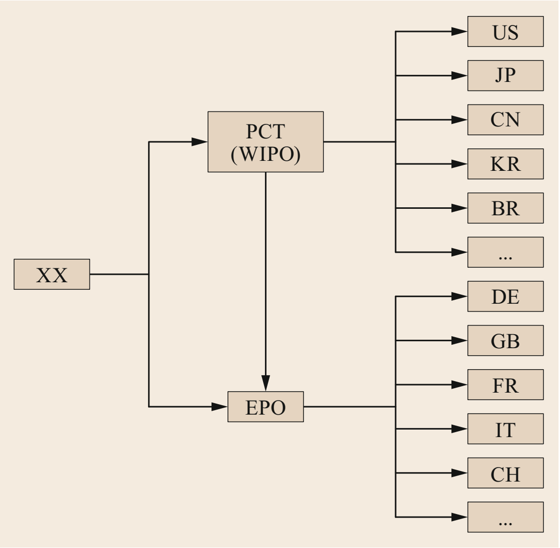


Fig. 37.5 Illustration of procedural paths of international and European patent applications

For example, to seek patent protection in Europe, applicants can submit patent application in each national patent office where he wishes to protect his invention (direct route; this direct route is not illustrated in Fig. 37.5), or he can make a single EPO patent application (EPO route), designating EPC member states, or file a PCT international application (PCT route). Figure 37.5 illustrates the use of the EPO and the PCT systems to seek patent protection in multiple countries.

When counting patent applications for statistical analyses, these structures have to be taken into account. Thus, when counting applications from foreign countries in a specific destination country, all possible ways of applying must be considered:

- Direct applications
- Applications through the EPO
- Applications through the PCT.

All patent applications referring to the same priority application/invention constitute a so-called patent family. Some databases display patent families, so that it is possible to count unique inventions and to assess the international marketing intentions linked to a patent application.

The country codes in Fig. 37.5 follow the alpha-2-code of the ISO 3166-1 standard (https://en.wikipedia.org/wiki/ISO_3166-1) and are also used in patent documents and databases.

37.5 International Country Comparisons

The information on international patent flows, international application paths, and application costs are a necessary basis to conceive appropriate concepts for country comparisons on the basis of patent indicators. At present, three approaches are suggested:

- Patent families with at least two members
- The IP5 approach
- Transnational patents.

A frequently made proposal is to use applications in foreign countries, as these applications imply much higher costs than domestic applications, hence capturing the most valuable inventions. Formally, this means that only patent families (patent families are defined in Sect. 37.4) with at least two members counted, meaning patent families with a domestic application and at least one application in a foreign country. The drawback of this approach is that—due to the specific geopolitical situation—the propensity to patent in foreign countries differs between countries of origin. For instance, the step from the Netherlands to Germany implies a family size of two members, as does the step from China to the US. Counting only families with at least three members improves international comparability and captures only high value patents, but still does not lead to a satisfying outcome. Counting on the basis of family size is often used in official statistics, whereas it is less available in online databases. The family size can be determined via an in-house database such as PATSTAT as well. A disadvantage of family-based searches is, of course, that the sample sizes of analyzed patent applications become much smaller. Thus, for more detailed analyses on the country level without a country comparison, the use of domestic applications is still helpful.

Against the background of the current international structures, the OECD suggested the *IP5* approach where patent applications to the five major patent offices in the world are considered [37.10, p. 20]:

1. The EPO
2. The USPTO
3. The JPO
4. The SIPO
5. The KIPO.

The OECD suggests three possible versions of this approach [37.10, p. 20f]:

1. Families of patent applications with members filed with one or more IP5 offices, including single filings. This implies that applications filed only with one of the IP5 offices, i. e., the EPO, the USPTO, the JPO, KIPO and SIPO, are considered.
2. Families of patent applications with members filed at least with one of the IP5, excluding single filings. This implies that applications filed only with one of the IP5 offices ... are considered only in so far as another family member has been filed with any other office worldwide (anywhere in the world, not necessarily with another IP5 office).
3. The most restrictive definition is that families of patent applications are considered only in so far as family members have been filed with at least two IP5 offices. For instance, patents filed with the USPTO will be considered only if an equivalent filing has been made with at least one of the remaining four IP5 offices. This is irrespective of whether equivalent applications in non-IP5 offices also exist.

The first option leads to considerable distortions between countries due to the dominant weight of domestic offices. According to the OECD, the second option leads to appropriate results, and the outcome of options 2 and 3 are very similar [37.10, p. 22]. The major shortcoming of the second approach for practical uses is that it is quite difficult to compute, but it is feasible.

The third approach is easy to compute. It could, however, be that the threshold for Southeast Asian countries is too low, as, e. g., the step from China to Japan is smaller than that from Europe to the US. Furthermore, the approach will yield appropriate data for current years only if PCT applications aimed at the IP5 offices are included. Due to the substantial delay of the transfer from the international PCT phase to the destination offices, these data appear quite outdated.

An alternative to balanced country comparisons is so-called transnational patents. These are patent applications either at the EPO or PCT applications without double counting [37.11]. In other words, these are patent families with at least an application at the EPO or the WIPO via the PCT route. It may be irritating that EPO and PCT applications are treated as equivalent, but the outcomes lead to useful results, as documented in *Frietsch* and *Schmoch* [37.11]. The major advantages of this approach are the generation of quite large samples, the limited delay of 1.5 years after the priority date, and the easy computing of searches.

Figure 37.6 shows a comparison of the IP5, the family ≥ 2 and the transnational approach, for the field of robotics, defined by B25J/IPC and `robot*` as keyword (with open truncation). The IP5 and transnational trends are similar, but not identical, and the absolute level of the IP5 approach is nearly identical to the transnational one. In the most recent years, the curve of the IP5 approach has been sloping downward, as the PCT applications not yet transferred to one of the IP5 offices were not included. The curve based on families with at least two members (foreign-oriented families) is lower than the two others in absolute terms, but the trend is similar. Yet again, one can observe a downward slope in the most recent years.

The authors of the transnational approach conceived it at a time when the growth of PCT applications was visible but less pronounced than currently. In 2000, 57% of the EPO applications were Euro-PCT applications; in 2012 this share was 72%. It is possible that with an increasing use of Euro-PCT applications and an increasing relevance of Southeast Asia as an economic region, the exclusive analysis of PCT applications may be more appropriate for balanced country comparisons than the transnational approach. The situation must be checked regularly.

To conclude this comparison, the necessity to specify the types of patent applications included is illustrated in Fig. 37.7. Yet again, the relationship between the five major countries is shown for a search without restrictions for the example of robotics according to the approach of transnational patents, the IP5 approach (according to version 3) and the family ≥ 2 approach:

- Without further specification, China seems to dominate the patent applications in robotics by far, a statement which is not realistic at all, according to experts in the field.
- Following the transnational concept, the US is the leading country, and Japan achieves a comparable level.
- In the IP5 approach, Japan is the leading country. In the IP5 approach, the position of Germany appears to be quite far behind, particularly compared to China and South Korea, a result that does not fit with other economic data on robotics. The IP5

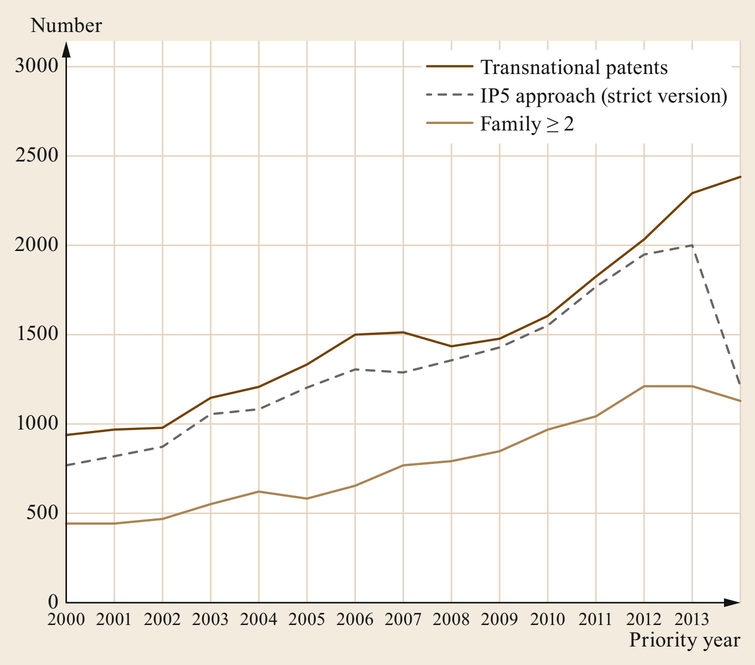


Fig. 37.6 Patent applications in robotics, according to the different approaches (after [37.3, 12])

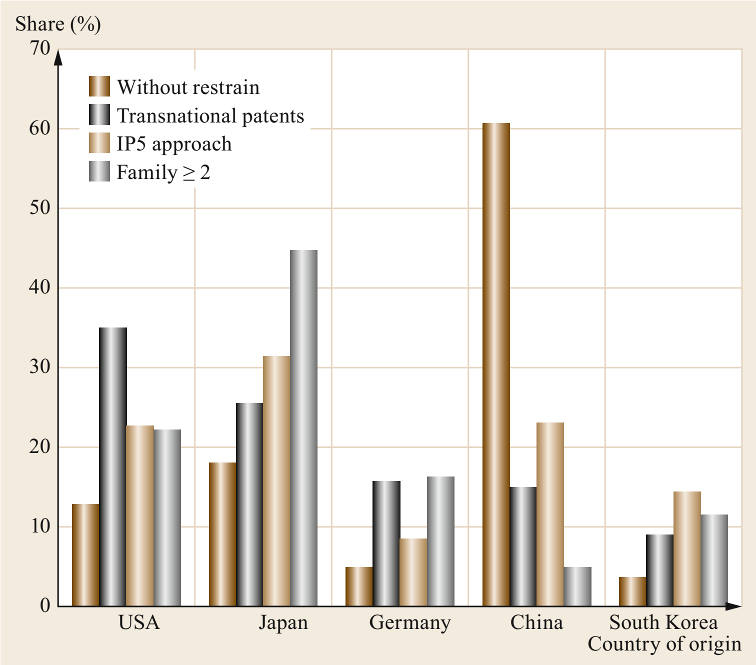


Fig. 37.7 Share of patent applications in robotics for selected countries, according to different counting approaches, 2014 (after [37.3])

approach seems to be better than that, an approach without restraints, but less realistic than the transnational one.

- In the family ≥ 2 approach, Japan earns a higher position than the US, and the position of China is very far behind. Again, the latter approach seems to be less appropriate for country comparisons than the transnational one.

This is only a conclusion for a specific example. A more detailed analysis is necessary. Anyhow, *Frietsch* and *Schmoch* [37.11] were able to show that the relative values between countries in the case of transnational patents highly correlate with input data (R&D shares) and output data on international markets (foreign trade). In any case, a well-considered approach of country comparisons is crucial for appropriate patent statistics.

37.6 Effectiveness of Keyword Searches

A further aspect implying misleading outcomes is the overestimation of the utility of keyword searches. In most cases, the search by codes of the International Patent Classification (IPC) is sufficient, as this scheme encompasses about 70 000 codes. However, in some cases, especially in very new areas, searches by keywords are necessary to complement the IPC searches. Here it is often overlooked that in the European patent systems the legal requirements for patent abstracts are quite low and that in some cases the applicants even try to hide the content of their invention to misinform competitors. The consequences are illustrated by the example of graphene, a new two-dimensional (2-D) material. The example was selected, as in an invention referring to graphene, it seems to be nearly impossible to avoid the keyword `graphene`, thus it is a strong keyword. Looking at the official standard abstracts of applications at the EPO, about 340 applications are found in 2013. For example, in the database PATSTAT, only the official abstracts are documented. The database World Patents Index (WPI) offers enhanced abstracts and titles focused on the technical content of the inventions. With this database, about 700 applications, thus more than double, can be identified (Fig. 37.8). Looking at the full text of the applications, 740 applications are discovered, but it is not clear whether the invention refers to graphene, whether it is compared to graphene, or what other reasons for using `graphene` apply.

In the case of the USPTO, the number of graphene applications identified by standard and enhanced abstracts is equivalent, which documents the high legal requirements for the content of patent abstracts

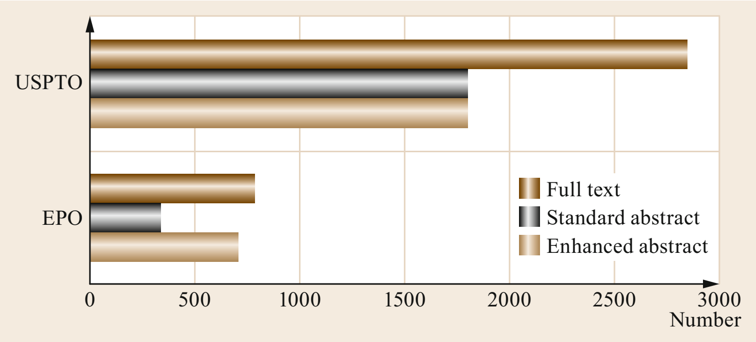


Fig. 37.8 Number of patent applications in graphene at the EPO and the USPTO based on keyword searches, priority year 2013 (after [37.3, 13, 14])

according to the US patent law. Again, the search in the full texts provides more results, but with some uncertainties. The absolute differences between the results for the EPO and the USPTO are primarily due to the fact that the USPTO is a national office and covers many domestic applications, whereas the EPO is a *regional* office aiming at foreign applications.

37.7 Features of the Cooperative Patent Classification

A regular problem of patent searches is the classification of European patents and most other patents worldwide by the IPC and the classification of US patents by the US patent classification (USPC). For many years, the US documents have been also labeled by IPC codes, but the philosophy of the IPC and USPC often differs, and the automatic concordance between IPC and USPC can be misleading. This is not only a problem for scientists, but also for patent examiners in the case of novelty searches. Against this background, the EPO and the USPTO launched the initiative "of developing a transparent and harmonized approach to a global classification system for patent documents" [37.15] in 2010, the Cooperative Patent Classification (CPC). The CPC is already implemented in databases, and the question arises as to whether a search with IPC codes or with CPC codes is more appropriate.

The available experience shows that in most cases the results of IPC and CPC searches are equivalent. Only in new fields, do some differences come up, as illustrated by the example of nanotechnology classified in the IPC as well as the CPC in the class B82. In recent years, the results for the CPC and IPC searches have been almost identical, but between 2000 and 2006 the numbers for the CPC search are substantially higher than for the IPC search (Fig. 37.9). As no systematic reason for this difference can be found, the only solution is to use both classifications in parallel, in particular in the case of new, emergent technologies.

In general, the introduction of the CPC implies considerable advantages for patent searches, so it is explained in more detail.

The CPC was initiated as a partnership between the USPTO and the EPO, where the offices agreed to harmonize their existing classification systems (ECLA and USPC, respectively) and migrate towards a common classification scheme. This was a strategic decision by both offices and is seen as an important step towards advancing harmonization. At the USPTO, the conversion will provide an up-to-date classification system that is internationally compatible (http://www.cooperativepatentclassification.org/about.html).

The migration to CPC was developed based in large part on the existing European classification (ECLA) system modified to ensure compliance with the international patent classification system (IPC) standards administered by the world intellectual property organization (WIPO). The previous European classification system (ECLA) was a more specific and detailed version of the IPC system. The following sections are identical in both the CPC and the IPC:

- A: Human necessities
- B: Performing operations, transporting
- C: Chemistry, metallurgy
- D: Textiles, paper
- E: Fixed constructions
- F: Mechanical engineering, lighting, heating, weapons
- G: Physics
- H: Electricity

In addition, the CPC encompasses:

- Y: Emerging cross-sectional technologies

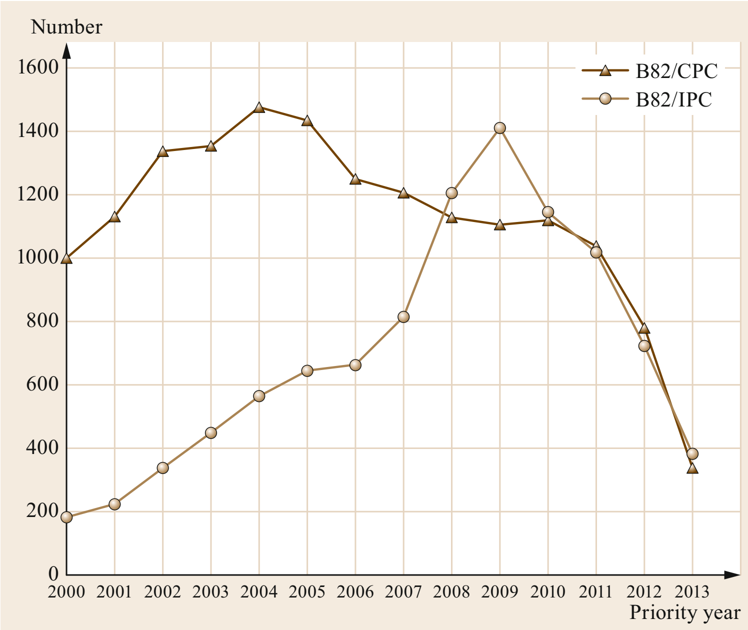


Fig. 37.9 Patent applications at the EPO in nanotechnology (after [37.3])

This new section covers:

- "Technologies or applications for mitigation or adaptation against climate change"
- "Information or communication technologies having an impact on other technology areas"
- "Technical subjects covered by the former USPC".

A detailed list of the presently valid CPC codes is available on a joint website of the USPTO and EPO [37.16, 17].

The CPC will have more subgroups than the IPC (about 70 000) and the ECLA (about 140 000 subgroups) and will achieve a level of about 400 000 subgroups, thus the CPC will provide a very refined classification.

To illustrate the substantial advantages of the CPC, the example of the subgroup B25J 13/086 is shown in Fig. 37.10. Compared to the IPC, nine further subgroups are added specifying the sensing devices (Fig. 37.11). Furthermore, the classification scheme is complemented by a detailed definition of the codes aiming at an internationally harmonized understanding of how technological objects should be classified appropriately. For example, the definition of the main groups and subgroups of the subclass B25J is a document of 143 pages.

In a joint effort, the EPO and USPTO classify all documents with CPC codes from 2013 onwards. According to the CPC annual report of 2015, more than 97% of US documents are classified with CPC codes, as are almost 100% of the documents at the EPO and

13/00 Controls for manipulators (programme controls B25J 9/16; control in general G05)
…
13/08 . by means of sensing devices, e.g. viewing or touching devices
…
13/086 .. {Proximity sensors}

Fig. 37.10 Cut out of the subclass B25J of the CPC (status October 2016)

13/00 Controls for manipulators (programme controls B25J 9/16; control in general G05)
…..
13/08 **.** by means of sensing devices, e.g. viewing or *touching devices*
13/081 **. .** {Touching devices, e.g. pressure-sensitive}
13/082 **. . .** {Grasping-force detectors (in general G01L 5/16, G01L 5/22)}
13/083 **. . . .** {fitted with slippage detectors}
13/084 **. . .** {Tactile sensors (in general G01L 5/16, G01L 5/22)}
13/085 **. .** {Force or torque sensors (B25J 13/082, B25J 13/084 take precedence)}
13/086 **. .** {Proximity sensors}
13/087 **. .** {for sensing other physical parameters, e.g. electrical or chemical properties}
13/088 **. .** {with position, velocity or acceleration sensors}
13/089 **. . .** {Determining the position of the robot with

Fig. 37.11 Scheme of subclass B25J 13/086 of the CPC (after [37.16])

B25J 13/086
{Proximity sensors}
Definition statement
This place covers:
Illustrative example of subject matter classified in
B25J 13/086 EP 0518836

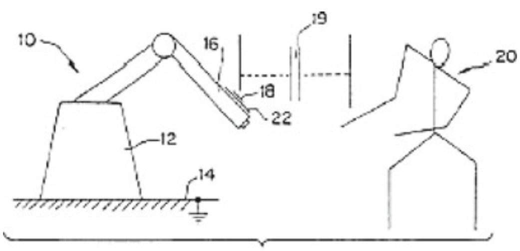


FIG. 1

Fig. 37.12 Definition of subclass B25J 131086 of the CPC

WIPO [37.18, p. 7]. For patent statistics, this means that the documents of the large western offices can be searched via CPC with only very few documents missing. In the case of China and South Korea, the coverage is still incomplete, so that the searches should be performed via IPC codes. For documents created before 2013, searches have to be conducted on the basis of IPC or USPC characters. Thus, for longer time series, the old and new systems must be used in parallel. Of course, this is only a problem of the present transition period. In the long run, the CPC will be very advantageous.

In practice, the CPC will be very helpful for using the database PATSTAT, which covers the IPC as well as the CPC, but which is less appropriate for key word searches, as illustrated above. For identifying the appropriate CPC code, it is recommended to start by looking for appropriate IPC codes in standard databases, e. g., in the WIPO file of IPC codes (http://www.wipo.int/classifications/ipc/en/). In the next step, the CPC scheme and definitions can be checked for finer codes of the CPC. Access to the scheme and definitions are provided by the CPC website [37.16]. Figure 37.11 demonstrates the scheme of the subgroup B25J 13/086 for which Fig. 37.12 shows the definition in an illustrative example.

If finer, more specific CPC codes can be identified, they may replace keyword searches and substantially improve the results in databases such as PATSTAT, which exclusively cover the official text elements of patent applications. However, there are still cases where keyword searches cannot be avoided, e. g., in cases of very new emerging fields that are not yet covered by IPC or CPC codes. Furthermore, generic technologies with a broad variety of applications exist in different fields. In these cases, the referring patent applications are generally classified either in the core area of the technology or associated technology areas or areas of application, and there is no common code covering the core technology, associated technologies, and all application areas as a common bracket.

An example for such a situation is robotics. The core technology is classified in the subclass B25J. Associated technologies are measuring, control, or computing. Application areas are machine tools, painting, cleaning, agricultural machines, surgery, etc. For patent searches, this implies that not all patent applications referring to robotics are classified in B25J. For example, a combination of the keyword `robot` with right-hand open truncation and the IPC subclass B25J leads to 2405 transnational patent applications in the priority year 2014, if the keyword search is performed in enhanced abstracts as in the WPI database. In this dataset, only 46% of the documents are identified by the IPC code. This share is even lower, when the search is conductd by CPC codes, if codes of the CPC subclass Y10S (technical subjects covered by former USPC cross-reference art collections (XRACs)) are used in addition to B25J.

Due to the relevance of keywords it is not advisable to perform searches with PATSTAT in these types of cases, as PATSTAT only provides official titles and abstracts, not enhanced abstracts like the WPI. A possible alternative is to identify all publication numbers of patent applications linked to a specific topic, e. g., to robotics, by statistical analysis, to generate a table of publication numbers for robotics in PATSTAT, and to conduct all other analyses such as trend, country, applicant, or subfield analyses in PATSTAT.

A further problem in the context of classifications is the frequent update of classifications. This has been a standard at the USPTO for many years, but until 2005 the IPC was updated every 5 years only. This frequent update means that former patent searches based on IPC codes can become invalid. For instance, in 2016, the WIPO found that a classification of biotechnology of 2005 had to be changed at about 60 positions, where some groups were moved to other subclasses or new codes were introduced. This change in update requires new practices for users of IPC codes. In particular, if searches for some fields are regularly updated, the validity of the IPC definition has to be checked each time.

Patent analyses have found broad propagation in studies for official authorities such as ministries, political commissions, or governmental institutions. This broad acceptance of patent analyses can be seen as a positive trend, however, the expectations of what can be achieved are often too high. A major problem is that politicians are confronted with fuzzy buzzwords and hope that patent analyses can clarify the situation. Typical buzzwords are *internet of things*, *advanced materials*, *advanced computing*, or *photonics*. The problem in these cases is that patent analyses need clear technological definitions. Thus, e. g., the question must be answered as to which types of material may be considered as advanced, polymers, metals, or ceramics with special properties, nanomaterials, special coatings, etc. At the beginning of the analysis, someone must take the decision which materials are advanced ones. A frequent solution is to collect many IPC codes of materials which may be considered as advanced leading to a fuzzy mix of different materials/technologies and thus avoiding a clear decision as to which technologies are relevant. The outcome will be fuzzy as well and is not interpretable in a clear way. Here, we run into typical problems of composite indicators, where the reasons for a positive or negative result cannot be identified [37.19]. Thus, it would be much better to identify ex ante precisely which items are representative for a field.

37.8 Patents of Large Companies

Patent databases represent large, detailed samples of indicators on innovation activity and are a popular source for many types of innovation analysis. The incontestable advantages often seduce scholars into neglecting the shortcomings in patent statistics. As early as 1985, *Pavitt* [37.20] carefully investigated possible problems of patent statistics and enumerated three sources of bias:

1. differences amongst countries . . .
2. differences amongst technologies and sectors . . .
3. differences amongst firms . . . [37.20, p.82].

In the first part of this chapter, we already addressed the typical problem of differences amongst countries. In the case of technologies and sectors, the differences in the propensity to patents are often overlooked. A typical example of different propensities to patents in technologies is the relatively high patent numbers in solar cells versus moderate ones in wind energy. Thus, a direct comparison would lead to higher patent numbers for solar cells, which does not mean that the economic value of solar cells is higher than that of wind mills, but they could be less technology intensive and/or imply more mature technologies, implying less recent patent activity.

In these cases, it is advisable to refer to other sources such as turn-over, R&D, etc., as a reference. As far as firms are concerned, it must be taken into account that the propensity to patent differs even between firms in the same sector or technology. Thus, a higher number of patents is only a strong hint at substantial innovation activities of a firm, not a proof. A systematic bias in patent statistics is the underrepresentation of small and medium-sized enterprises (SMEs). In their perspective, the costs of patent protection in relation to their markets are relatively high and in the case of infringement they have moderate chances to successfully enforce their rights against large firms due to considerable legal costs. Therefore, SMEs often prefer other strategies such as trade secrets or speed to the market [37.21]. Nevertheless, the contribution of SMEs to new technology is considerable, as documented, e. g., in Eurostat 2014. It is necessary to have these limitations in mind, and in some cases, it can be useful to complement patent analysis by surveys to get an impression of the size of the bias.

A further problem of the statistical analysis of enterprises is the appropriate count in applicant lists. In ranking lists of applicants, it is possible that some enterprises appear in different name variants, so that it is necessary to unite these different versions to achieve an appropriate list. To avoid individual manual cleaning, *Magerman* et al. [37.22] developed an automated method of name harmonization with quite advanced methods. However, this approach cannot solve the problem that many large enterprises have affiliations with completely different names, e. g., the affiliation Genentech of Hofmann La Roche would not be identified as part of Hoffmann La Roche by automatic approaches. *Magerman* et al. [37.22, p. 4] correctly state:

> when harmonizing legal entities, every patentee name needs to be checked against historical information on naming practices and ownership to address the following issues:
>
> - Identification of entities (business units, departments, subsidiaries) that may have a different name but belong to the same legal entity
> - Identification of name changes over time
> - Identification of mergers and acquisitions
> - Identification of joint ventures
> - Identification of mother and daughter relationships/subsidiary companies.

Even, if these problems are solved, it is not always possible to add up the patent numbers of different parts of the same entity in applicant lists. In particular in PCT applications, aiming at different countries of protection, it is possible that different subentities of an enterprise appear on the same patent application, for instance, PCT applications exist with two assignees of the same legal entity, e. g., Hoffmann La Roche Inc. (USA) and Hoffmann La Roche & Co AG (Switzerland) or Hoffmann La Roche Inc. and Genentech. Here, the applicants in the final list cannot be added up, but some must be deleted. Thus, also in patent statistics, problems with different name variants similar to bibliometrics exist, but for different reasons.

A relevant recent trend of patent statistics is to analyze enterprises instead of countries with the argument that country boundaries lose relevance due to the high activities of multinational enterprises. However, such enterprise-oriented analyses run into the methodological problems addressed above. The first problem is the appropriate definition of enterprises with all their affiliations and name variants, the second problem the comparison in an international context.

Ranking tables of large international enterprises are provided by the OECD and the WIPO. The OECD invests enormous efforts in the definition of enterprises [37.10, p. 16]. Nevertheless, *Neuhäusler* et al. [37.23] find different results by an automatic approach linked to enterprise databases. Furthermore, it is important to be aware that the company rankings differ depending on the criterion of patent counting. The OECD uses the definitions 2 and 3 of the IP5 approach.

A further provider of company rankings is the WIPO. It rates the patent applications of the companies by the number of families [37.24, p. 11ff]. However, the distribution of the number of family members differs considerably by company, so that a simple ranking is difficult.

To demonstrate the effect of different rating methods of patent applications, we selected 19 large companies that appear in the lists of the OECD and the WIPO. When simply the number of patent families is counted, a list according to Table 37.5 is generated. The analysis is based on simple name searches, so that affiliations with different names are not included. The result is, therefore, not exact. When the same search is based on transnational patents, thus patent families with higher economic value, a quite different list appears (Table 37.6). In Table 37.5, Southeast Asian companies dominate, as also families with one member are included. The effect of the *low average value* of patent applications at Southeast Asian offices as illustrated in Table 37.1 becomes visible. When only patent families with a higher economic value are included, European and US American companies achieve higher positions. To appropriately assess such rankings, it is also necessary to consider the main areas of activity of the companies considered and to standardize the patent numbers with field averages, as the propensity to patent varies by field. To conclude, company rankings based on patents are methodologically challenging and are difficult to interpret.

It may be argued that such lists of very large enterprises do not adequately reflect economic reality. As the economic prosperity of countries depends on the activities of large enterprises as well as SMEs, which

Table 37.5 Ranking of selected companies by the number of patent families, 2014 (after [37.3, 5])

Company	Patent family
Samsung Electronics	14 902
LG Electronics	10 844
Mitsubishi Electric	10 026
Hitachi Ltd.	9537
Canon	8672
International Business Machines (IBM)	7319
Toyota Jidosha	7246
Seiko Epson	6597
Toshiba	6459
Panasonic Corporation	5502
Robert Bosch GmbH	4418
Siemens AG	4186
Ricoh	4036
Fujitsu Ltd.	3996
Sharp	3215
Denso	3118
Sony Corporation	2743
Hon Hai Precision	2231
General Electric Company	1997

Table 37.6 Ranking of selected companies by the number of transnational patent applications, 2014 (after [37.3, 5])

Company	Patent applications
LG Electronics	3476
Samsung Electronics	3202
Mitsubishi Electric	3101
Hitachi Ltd.	2937
Siemens AG	2424
Sony Corporation	1955
Panasonic Corporation	1685
Robert Bosch GmbH	1671
Sharp	1435
Toyota Jidosha	1418
Denso	1206
Toshiba	1016
Seiko Epson	865
General Electric	792
Canon	712
Fujitsu Ltd.	691
Ricoh	441
International Business Machines (IBM)	142
Hon Hai Precision	25

contribute to a considerable extent to economic welfare in terms of employment, turnover, etc., or as suppliers or clients of the very large enterprises, it is useful to complement the data for very large enterprises by those of large enterprises and SMEs. For instance, *Frietsch* et al. [37.25] matched the names of applicants from nine different countries with the enterprise names in enterprise reference manuals such as Amadeus. On this basis, it is possible to examine the relevance of large enterprises and SMEs for a country in more detail [37.26]. The possibility of linking such databases to patent data is a considerable advantage of PATSTAT.

37.9 Patent Value

Standard patent statistics assume that patent applications have an equal or at least similar value by counting the number of patent applications. A relevant step towards a more realistic view is the analysis of IP5 applications, the family > 2 or transnational patents, thereby focusing on more valuable patents, i. e., patents targeting foreign market, thus approaching this assumption of similar value. Nevertheless, it is interesting to look at the question of patent value in more detail.

In principle, different types of patent value may be considered: the technological, the economic, the social or the strategic value (a more detailed discussion of the different forms of patent value can be found in *Frietsch* et al. [37.27, p. 10ff]). In the context of this chapter, the focus lies on the economic value.

It is important to be aware that the technological value and the economic one are not identical. For instance, it may be that a measuring device for scientific laboratories is extremely complex and has a high technological value, but the corresponding market is quite small, so that the economic value is moderate. Or vice versa, a technologically simple household appliance may have a broad market and a high economic value.

Several studies were conducted to assess the economic value of patents. A quite revealing one is that of *Gambardella* et al. [37.28] based on a survey of at inventors with referring to about 8200 granted patents at the EPO. Without discussing methodological details, we directly present the main result of this survey, the distribution of values in this large sample. Figure 37.13 shows that the patent value distribution is skewed (see also [37.29]). Since the difference in the logs of the boundaries of the intervals is roughly constant, the distribution in the figure approximates a log-normal. Even the log-normal distribution looks skewed. The estimated mean of the patent value distribution is higher than 3 million Euros and the median is about 400 000 Euro.

This skewed distribution raises the question what an appropriate basis of patent analysis, e. g., for country comparisons, is. For this purpose, we used the data of *Gambardella* et al. [37.28] to estimate the distribution of total values linked to the distribution of Fig. 37.13. We multiplied the mean value of each value class with the number of answers of this class which leads to a distribution of the total values according to Fig. 37.14. Again the distribution of the total values is quite skewed. The highest value class, representing about 1% of all patents, collates about 70% of the total value of all patents in the sample.

As the sample of Gambardella et al. refers to patents granted at the EPO, all these patents are transnational ones, defined as either EPO or PCT applications without double counting. So even in the case of transnational patents a relatively high share of the patents has a moderate value. e. g., about 40% of the patents in Fig. 37.14 have a value below 300 000 Euro.

This analysis raises the question what issue is analyzed by patents. If we use an approach such as IP5, family ≥ 2 or transnational, patents of very low value are excluded. Using the transnational approach, about 40% of the applications of European inventors which are applied exclusively at the domestic office, are excluded. For applications at the EPO, it can be assumed that the applicants expected a potential high value. Patent applications are always risky and only a small share will generate a high economic success. Thus, it is probable that among the 40% of the patents of lower value are many intermediate inventions within a longer technology line, where only the final invention implies considerable economic success. Thus, by the analysis of patents, we observe a broad variety of technological innovation activities including precursors of successful results.

In the case of the right end of the value distribution, most of the applicants will be multinational enterprises (MNEs) which can exploit their inventions in many countries at a world-wide level. As the production of the associated goods will take place in various national affiliations of these MNEs, the inventions will only partly be linked to foreign trade. Of cause, it is possible to analyze such types of inventions by introducing high family sizes of more than 6, but the final question is which type of activities should be examined: activities of MNEs, of SMEs, national comparisons, enterprise comparisons, etc.

Anyhow, *Frietsch* and *Schmoch* [37.11] were able to show that the relative values between countries in the

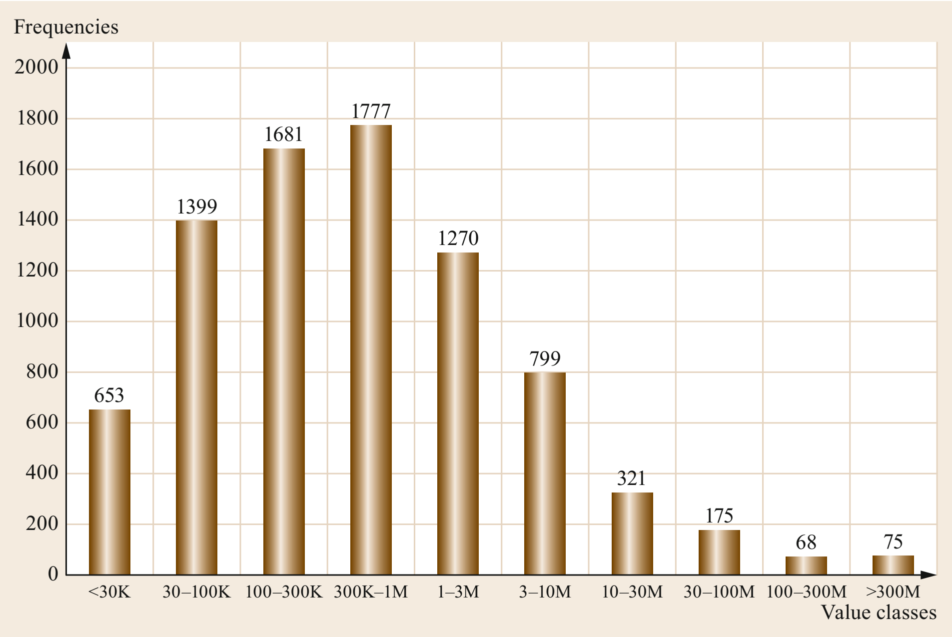


Fig. 37.13 Distribution of patent values (after [37.30], with permission from Wiley)

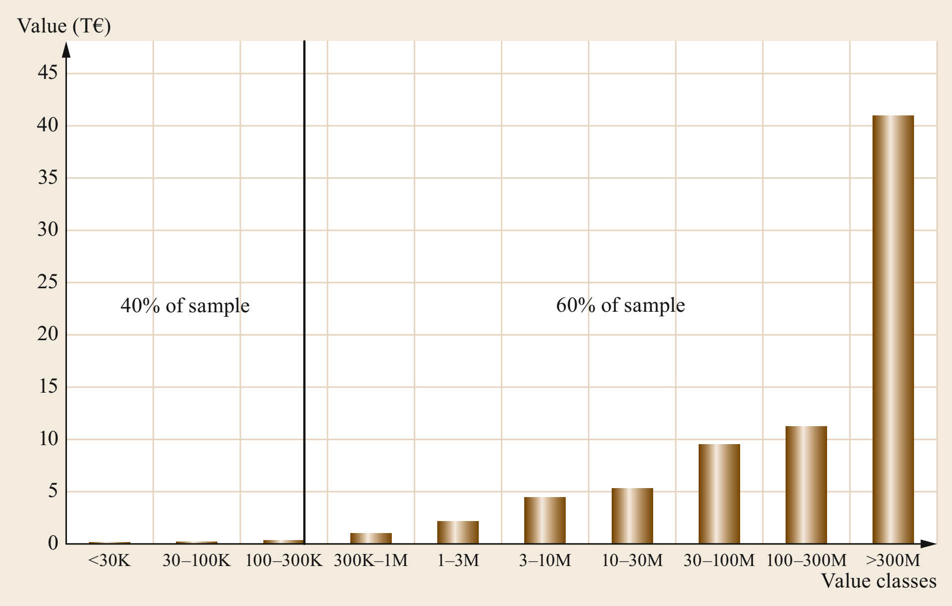


Fig. 37.14 Distribution of total values of patents in the survey of *Gambardella* et al. (after [37.28])

case of transnational patents—thus without focussing on patents with extremely high value—highly correlate with input data (R&D shares) and output data on international markets (foreign trade). *Frietsch* et al. [37.27] tried with a variety of different approaches to improve country comparisons with a more distinct focus on high value patents, but found no convincing solutions. However, in the case of very large enterprises such as shown in Sect. 37.8, it seems to be appropriate to limit the analysis to patents with very high value, e. g., by limiting the analysis to, e. g., more than 6 family members.

Several indicators have been suggested for measuring the economic value of patents.

A primary indicator in this context is forward citations of patents. This indicator was suggested quite early on by *Narin* and *Noma* [37.31] and validated by later studies, particularly by *Harhoff* et al. [37.32]. For this purpose, patent holders were asked to assess the asset value of their patent rights. The survey considered all patent grants with a 1977 German priority date which were renewed to full term, thus a sample of valuable patents. In the latter study, various potential indicators of patent value were compared, and the authors identify forward citations as a strong indicator for patent value. This finding is confirmed by various other studies, e. g., by *Hall* et al. [37.33]. In a later study, *Gambardella* et al. [37.28] confirm that forward citations are the best indicator for patent value, but they find that they explain only a small share of the variation in the patent value.

The major shortcoming of this indicator is the time lag between the submission of an application (priority date) and the possibility of conducting sound assessments. In the case of EPO applications, a time window of at least four years is recommended [37.27]. A fixed citation window is necessary to compare enterprises or countries, as the citation rate increases with a greater time lag. This indicator proves to be useful for individual cases, country comparisons are feasible, but require a complex methodology [37.27, p. 44 ff].

Family size is another indicator of patent value with the advantage that the data are easy to search and are available with the publication of the applications, i. e., 18 months after the priority date. This indicator is useful for comparing the patent value of enterprises within one country. However, the comparison of countries is problematic due to the different geopolitical positions of countries as has already been discussed in paragraph 5. For international comparisons it is recommended to exclusively consider large family sizes of more than 4 family members to compensate the geopolitical bias. In any case, family size appears to be the best criterion to select patent applications with a very high value. A further advantage is the availability at a very early point in time briefly after the priority application.

A further relevant indicator of patent value is opposition. At the EPO, third parties may dispute the validity of a granted patent by filing an opposition 9 months after the publication of the granting decision. As the opposition procedure is time and money consuming, it makes sense to assume that opposition is only raised in cases of valuable patents. The close relation of opposition and value has been shown several times. The disadvantage of this indicator is the considerable time lag between priority date and opposition, the quite small sample sizes—opposition is a rare event—, and the limitation on a simple valuable/less valuable decision; therefore value levels cannot be defined.

In the case of the US law, a post-granting review similar to the EPO opposition is possible under the new patent law of 2012. This relatively new provision has not yet been examined as the link to patent value, but it can be assumed that similar positive validations will be observed.

For litigations, similar considerations as for opposition apply. If litigations occur, a high value of the patent concerned is probable, but the event is quite rare, too. The access to litigation data via databases is difficult, but feasible. There are some studies on this issue, e. g., *Cremers* et al. [37.34] or *Lanjouw* and *Schankerman* [37.35]. Furthermore, many controversial cases are regulated between the interested parties outside of court, as in this area, lawsuits are very expensive.

Another possible indicator of patent value is renewals, assuming that the progressively increasing renewal fees reflect a substantive economic interest in a patent and thus its high value. This indicator is especially interesting regarding the USPTO with its high renewal fees after 10 years. Even if the renewal fees are only a marginal part of the total costs including the other fees and the costs of attorneys, the analyses based on this indicator lead to useful results. The main disadvantages of this indicator are the quite complex analyses which need to be performed and the substantial time lag.

In many studies it has been shown that the size of inventor teams is correlated to patent value. However, the distribution of the sizes is quite uniform, so that it is not useful for distinguishing between specific patents.

The granting of a patent is a value indicator as well. But due to different practices in different patent systems it can only be used within one system. As in the case of opposition, only the distinction between valuable and less valuable is possible. Value levels cannot be defined. Another possible indicator for patent value might be the volume of licensing and portfolio transaction. However, this interesting information is available for spectacular individual cases, but not in a systematic way. Also, the information of some patent offices on licensing is incomplete, as only some applicants use this way of publication to reduce application fees. In most cases, licensing is secret.

All in all, many different indicators for patent values have been suggested. To decide whether a specific patent has a high value, it is recommended to check all these indicators.

Contrary to expectations, the fact that a patent has been granted does not prove to be an appropriate indicator for the economic patent value. Firstly, the examination process is focussed on the technological value and less on the economic one. Secondly, obviously the patent systems are so different that domestic applicants achieve higher granting rates than foreigners. Thirdly, the granting rates differ by field. Fourthly, the time lag between priority and granting date substantially differs by country [37.27, p. 40ff].

According to various surveys of inventors and managers, the distribution of patent values is highly skewed [37.29, 36]. This is not a special characteristic of patents, but can be found in citation distributions in bibliometrics, firm size distributions etc. For patent analyses this means that the measures for country comparisons should be used for cutting off at least the large number of patents with low value. In the remaining samples, a skewed distribution of values cannot be avoided completely. Nevertheless, it is possible to conduct country comparisons without weighting the absolute numbers by citations or other value indicators, as only minor impacts on rankings can be observed [37.27, p. 111]. A strict limitation of the analyses on patents with a very high value would imply that the searches are—de facto—limited to large enterprises and the bias of patent analysis in favour of large enterprises would be amplified.

All in all, a statement of *Gambardella* et al. [37.37] is still valid:

> The 'measure of our ignorance' in this field is still too high. This paper finds that new and better explorations of the determinants of the economic value of patents are an important and largely underdeveloped area for future research.

37.10 The Impact of Legal Changes on Statistics

For practitioners in patent statistics, it is useful to follow the changes in patent law which might influence the statistical outcome. In the case of the US patent law, many reforms were introduced by the America Invents Act (AIA) where the change from the first-to-invent principle to the first-to-file one is the most discussed change (a summary of the AIA is given in [37.38]). However, it can be assumed that the impact on patent statistics is limited. More influential for statistics was the change in 2001 where the possibility of a pre-grant publication was introduced. Since that time, it has been possible to observe new technological developments at the US market at an early stage [37.39].

Patent analyses for European countries are already quite complex, as in addition to applications at the domestic offices, those at the European Patent Office (EPO) and at the world intellectual property organization (WIPO) via the Patent Cooperation Treaty (PCT) must be considered as well. At present a new change is imminent: After many years of negotiations the representatives of the EU member states achieved a breakthrough agreement in 2012: The European Unitary Patent will soon guarantee supranational protection for inventions in 26 countries across Europe. However, only 10 states have ratified the agreement until mid-2016, thus the agreement has not yet achieved legal validity [37.40]. At present, it is difficult to predict to what extent the Unitary Patent will be used instead of the standard EPO patent and what the effects on patent statistics will be. In view of various sceptical voices as to the usefulness of the system [37.41] it seems to be likely that the use will be limited.

37.11 Conclusion

All in all, the technological progress in the provision of patent data offers a variety of new ways of patent analysis. Nevertheless, it is necessary to carefully comply with basic methodological principles to achieve appropriate results. In particular, it is important to focus the analysis on economically important patent applications.

References

37.1 S. Hinze, U. Schmoch: Opening the black box. Analytical approaches and their impact on the outcome of statistical patent analyses. In: *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies on S&T Systems*, ed. by W. Glänzel, H. Moed, U. Schmoch (Springer, Dordrecht 2005) pp. 215–235

37.2 World Intellectual Property Organization (WIPO): World Intellectual Property Indicators 2014, http://www.wipo.int/edocs/pubdocs/en/wipo_pub_941_2014.pdf (2014)

37.3 STN International: World Patents Index, http://www.stn-international.com/dif-dwpi.html

37.4 OECD: OECD Statistics, https://stats.oecd.org

37.5 Clarivate Analytics: Derwent World Patents Index, https://clarivare.com/products/derwent-world-patents-index

37.6 H. Dernis, M. Khan: *Triadic Patent Families*, OECD Science, Technology and Industry Working Papers, No. 2004/02 (OECD, Paris 2004)

37.7 B. van Pottelsberghe de la Potterie, D. François: The cost factor in patent systems, Working Papers CEB from ULB No 06-002.RS, Bruxelles (2006)

37.8 G. de Rassenfosse, B. van Pottelsberghe de la Potterie: The role of fees in patent systems: Theory and evidence, J. Econ. Surv. **27**(4), 676–716 (2013)

37.9 J. Bessen: Estimates of patent rents from firm market value, Res. Policy **38**(10), 1604–1616 (2009)

37.10 OECD: *World Corporate Top R&D Investors: Innovation and IP Bundles* (OECD, Paris 2015)

37.11 R. Frietsch, U. Schmoch: Transnational patents and international markets, Scientometrics **85**, 185–200 (2010)

37.12 PATSTAT: Explanation of file contents, https://www.epo.org/searching-for-patents/business/patstat.html#tab-1

37.13 STN International: European Patents Fulltext Database (EPFULL), http://www.stn-international.de/epfull.html

37.14 STN International: US Patent Fulltext Cluster (USPATFULL), http://www.support.cas.org/products/stn/uspatfull-anonst

37.15 Cooperative Patent Classification: http://www.cooperativepatentclassification.org/index.html (2010)

37.16 Cooperative Patent Classification: CPC scheme and CPC definitions, http://www.cooperativepatentclassification.org/cpcSchemeAndDefinitions/table.html

37.17 The Unites States Patent and Trademark Office (USPTO): CPC–Cooperative Patent Classification, https://www.uspto.gov/web/patents/classification/cpc.html

37.18 CPC: Explanation of the classification scheme, https://www.cooperativepatentclassification.org/index.html (2010)

37.19 H. Grupp, T. Schubert: Review and new evidence on composite innovation indicators for evaluating national performance, Res. Policy **39**(1), 67–78 (2010)

37.20 K. Pavitt: Patent statistics as indicators of innovative activities: Possibilities and problems, Scientometrics **7**(1/2), 77–99 (1985)

37.21 A. Leiponen, J. Byma: If you cannot block, you better run: Small firms, cooperative innovation, and appropriation strategies, Res. Policy **38**, 1478–1488 (2009)

37.22 T. Magerman, T.B. van Looy, B.X. Song: *Data Production Methods for Harmonized Patent Statistics: Patentee Name Harmonization*, K.U. Leuven FETEW MSI Research Report 0605 (K.U. Leuven, Leuven 2006)

37.23 P. Neuhäusler, R. Frietsch, N. Bethke, C. Mund: *Assessing Companies' Capability to Develop Advanced Manufacturing Technologies in Selected Industrial Sectors*, Final Report JRC/SVQ/2014/J.2/0009/NC (Fraunhofer Institute for Systems ans Innovation Research ISI, Karlsruhe 2015)

37.24 World Intellectual Property Organization (WIPO): *World Intellectual Property Indicators 2015*, Economics and Statistics Series (WIPO, Geneva 2015)

37.25 R. Frietsch, P. Neuhäusler, O. Rothengatter: *SME Patenting–An Empirical Analysis in Nine Countries*, Fraunhofer ISI Discussion Papers Innovation Systems and Policy Analysis No. 36 (Fraunhofer ISI, Karlsruhe 2013)

37.26 Eurostat: *Patent Statistics at Eurostat. Mapping the Contributions of SMEs in EU Patenting* (European Union, Luxembourg 2014)

37.27 R. Frietsch, U. Schmoch, B. Van Looy, J.P. Walsh, R. Devroede, M. Du Plessis, T. Jung, Y. Meng, P. Neuhäusler, B. Peeters, T. Schubert: *The Value and Indicator Function of Patents*, Report to the EFI (EFI, Berlin 2010), Accessible through http://www.e-fi.de/fileadmin/Studien/Studien_2010/15_2010_Patent_Value.pdf

37.28 A. Gambardella, D. Harhoff, B. Verspagen: The value of European patents, Eur. Manag. Rev. **5**(2), 69–84 (2008)

37.29 G. Silverberg, B. Verspagen: The size distribution of innovations revisited: An application of extreme value statistics to citation and value measures of patent significance, J. Econ. **139**(2), 318–339 (2007)

37.30 Gambardella et al.: (2018)

37.31 F. Narin, E. Noma: Patents as indicators of corporate technological strength, Res. Policy **16**, 143–155 (1987)

37.32 D. Harhoff, F.M. Scherer, K. Vopel: Citations, family size, opposition and the value of patent rights, Res. Policy **32**(8), 1343–1363 (2003)

37.33 B.H. Hall, A. Jaffe, M. Tratjenberg: Market value and patent citations, RAND J. Econ. **36**(1), 16–38 (2005)

37.34 K. Cremers, M. Ernicke, F. Gaessler, D. Harhoff, C. Helmers, L. McDonagh, P. Schliessler, N. van Zeebroeck: *Patent Litigation in Europe*, ZEW Discussion Paper No. 13-072 (ZEW, Mannheim 2013)

37.35 J.O. Lanjouw, M.M. Schankerman: Characteristics of patent litigation: A window on competition, RAND J. Econ. **32**(1), 129–151 (2001)

37.36 F.M. Scherer, D. Harhoff: Technology policy for a world of skew-distributed outcomes, Res. Policy **29**, 559–566 (2000)

37.37 A. Gambardella, D. Harhoff, B. Verspagen: The value of European patents, Eur. Manag. Rev. **5**(2), 69–84 (2010)

37.38 American Intellectual Property Law Association: IP policy and advocacy, http://www.aipla.org/advocacy/congress/aia/Pages/summary.aspx

37.39 U. Schmoch: Patent analyses in the changed legal regime of the US patent law since 2001, World Pat. Inf. **31**, 299–303 (2009)

37.40 Council of the European Union, The Member States: Agreement on a unified patent count, https://www.consilium.europa.eu/en/documents-publications/agreements-conventions/agreement/?aid=2013001 (2013)

37.41 R. Hilty, T. Jaeger, M. Lamping, H. Ullrich: *The Unitary Patent package: Twelve Reasons for Concern*, Research Paper No. 12-12 (Max Planck Institute for Intellectual Property & Competition Law, Munich 2012)

Ulrich Schmoch

Fraunhofer-Institut für System- und Innovationsforschung ISI
Karlsruhe, Germany
ulrich.schmoch@isi.fraunhofer.de

Professor Ulrich Schmoch holds a Doctoral degree in Social and Political Sciences (University of Hannover, Germany), a Diploma of Mechanical Engineering (University of Hannover) with modules at the University of Grenoble, France, and a State Doctorate in Sociology of Science and Technology (Karlsruhe Institute of Technology KIT, Germany). He has been working in different functions at the Fraunhofer Institute for Systems and Innovation Research (ISI), Karlsruhe, Germany, since 1986. Ulrich Schmoch has carried out and coordinated about 200 projects in the areas of innovation indicators, patent analysis, bibliometrics, systems of innovation, technology transfer, science management, innovation in threshold countries, innovations in the service sector, science-based technologies, and organization and mission of universities in an international comparison.

Mosahid Khan

World Intellectual Property Organization (WIPO)
Geneva, Switzerland
mosahid.khan@wipo.int

Mosahid Khan is head of Intellectual Property (IP) statistics at the World Intellectual Property Organization (WIPO). He is responsible for the collection and dissemination of IP statistics and actively involved in the development of methodologies and new IP-related indicators. He holds a master's degree in Economics from the University of London. Previously, Mr. Khan has worked as an economist/statistician at the OECD and his research work has been published in peer-reviewed journals.

38. Using Text Mining Algorithms for Patent Documents and Publications

Bart Van Looy, Tom Magerman

In this chapter we present an overview of text mining approaches that can be used to conduct science and technology studies that rely on assessing the (content) similarity between patent documents and/or scientific publications. We highlight the rationale behind vector space models, latent semantic analysis, and probabilistic topic models. In addition, several validation studies pertaining to patent documents and publications are presented. These studies reveal that choices in terms of algorithms, pre-processing, and calculation options have non-trivial consequences in terms of outcomes and their validity. As such, scholars should pay attention to the technicalities implied when engaging in text mining efforts in order for outcomes to become relevant and informative.

38.1 Text Mining and Science and Technology Studies

In this chapter, we discuss the nature and relevance of content analysis for conducting science and technology studies based on text mining algorithms. Text mining is already being used in efforts to delineate specific domains or subfields and/or to identify related developments in different activity realms. In the past, such demarcation relied heavily on existing classification schemes and expert opinions [38.1–8]. Current developments in the domain of text mining offer ample opportunities to improve such efforts by generating, on a large scale, automated results that indicate similarity between documents (e. g., patents, publications) and, hence, enable mapping, categorization, and classification.

Domain studies would not be alone in profiting from algorithms that enable the identification of con-

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_38

tent similarity across documents. Whether and to what extent knowledge spill-overs—between scientific and technological activity realms—are present could benefit from the ability to assess content-relatedness between patents and publications. A similar point can be raised with respect to the diversification strategies of firms. It has been argued that effective diversification benefits from the presence of knowledge spill-overs between existing and new activities. Whether, and under which conditions, this assumption holds, remains a point of scholarly debate, especially when firms are confronted with disruptive innovation. Future research efforts would benefit from being able to assess (dis-)similarities when firms simultaneously exploit existing capabilities and explore novel technological trajectories by analyzing the contents of the patent and publication portfolios of firms.

While the relevance of content analysis by means of text mining to conduct STI (science, technology and innovation) studies is clear, it remains unclear which set of text-mining algorithms will yield relevant (valid) insights when applied to data currently used in studies of science and technology (patent/publication titles and abstracts). In this chapter, we outline the principles and most commonly used algorithms related to text mining and assess their validity when conducting studies in the field of technology and science.

We begin this chapter with a brief outline of the nature of text mining (Sect. 38.1.1) and the origins of quantitative linguistics (Sect. 38.1.2), and follow this by outlining the different ingredients of text mining based on the vector space model (Sect. 38.2). Next, we discuss the rationale behind specific text mining models like latent semantic analysis and probabilistic topic models, and discuss similarity measures (Sect. 38.3). In Sect. 38.4, we present and discuss several validation studies that start from patent documents and publications. These studies reveal that, for science and technology studies, choices in terms of algorithms and calculation methods significantly affect the validity of the results obtained. In order to arrive at a comprehensive overview of relevant observations and insights, this chapter brings together a collection of insights and findings, which—to some extent—have been reported in [38.9–11]. Finally, the section on clustering and topic modeling was developed specifically for this handbook (Sect. 38.5).

38.1.1 Text Mining

Text mining refers to the automated extraction of information from text in order to reveal patterns that are present but not obvious in a document collection. Text mining can imply a broad set of activities, ranging from text categorization, text clustering, information extraction, sentiment analysis, document summarization, named entity recognition, and question answering; it is an interdisciplinary field based on natural language processing, computational linguistics, information retrieval, data mining, machine learning/artificial intelligence (cognitive science), mathematics, and statistics (for more information on the application of text mining and its relation to other fields and techniques, see [38.12–14], for an overview of techniques).

The large-scale availability of computing power has spurred considerable interest in text mining over recent years, together with the observation that the vast majority of (digitally available) information is stored as (unstructured) text.

Notice that text mining techniques go beyond mere information retrieval. Information retrieval helps in finding information based on a user request, and it is obvious that text mining techniques are instrumental in this respect. Indeed, currently information retrieval is probably the biggest application area of text mining techniques. However, information retrieval in itself does not reveal *new* knowledge or insights, it merely surfaces what is already assumed by somebody—the user issuing the search request knows what he/she is looking for (for an elaboration of this issue, see [38.12]). Text mining can go one step further by revealing patterns that were not obvious before. Compelling illustrations have been advanced by *Swanson*, where *unknown* relationships between fish oil and Raynaud's syndrome, migraine and magnesium, and somatomedin C and arginine have been advanced [38.15–17]. The relationship between *migraine* and *spreading depression* on the one hand, and *magnesium* and *preventing depression* on the other, was revealed after a thorough search of the medical literature on migraine, suggesting magnesium deficiency as a factor in migraine. Prior to this remarkable discovery—Swanson is an information scientist, not a physician—there was no indication of this relationship whatsoever; his results triggered additional clinical research, confirming his prognosis [38.18]. These case studies can be regarded as pioneering cases of text mining, providing the foundations for a more formalized study of literature-based discovery based on textual analysis—the so-called Swanson linking [38.19].

38.1.2 History of Quantitative Linguistics and Applications in Science and Technology Studies

The application of text mining techniques in science and technology studies is not new. Quantitative linguistics dates back to at least the middle of the nineteenth century [38.20]. The classical work by *Zipf* [38.21] is considered pioneering in quantitative linguistic (or text) analysis. Since the 1970s, a remarkable increase in activity has been witnessed in this area of information science. As for its application to scientific literature, *Wyllys's* study [38.22] is among the first. Co-word analysis, one of the most frequently used techniques, was founded on the idea that the co-occurrence of words describes the contents of documents and was developed for purposes of evaluating research [38.23]. The extension of co-word analysis to large sets of publications was possible as soon as large textual databases became available in electronic form; also, the increasing availability of computational power facilitated the further development and dissemination of text mining approaches. *Manning* and *Schütze* [38.24] provide a comprehensive introduction to the statistical analysis of natural language, *Berry* [38.25] provides a survey of text mining research, *Leopold* et al. [38.26] gave an overview of data and text mining fundamentals for science and technology research, and *Porter* and *Newman* [38.27] coined the term *tech mining* to refer to text mining collections of patent documents on a specific topic. Other earlier, practical applications in the field of bibliometrics and technometrics were offered by *Courtial* [38.28], *Noyons* et al. [38.5], *Bassecoulard* and *Zitt* [38.29], *Leydesdorff* [38.30], *Glenisson* et al. [38.3], and *Janssens* et al. [38.31].

As mentioned previously, a first and more traditional application of text mining is to be found in the field of information retrieval (conducting patent or publication searches on bibliographic databases). However, text mining offers a much broader range of applications:

- Domain studies: Starting from a set of seed patents or publications that are representative of a technological or scientific domain, concepts and topics can be extracted and used to match concepts and topics from other patents and publications with the purpose of identifying related patents and publications or in order to delineate technological or scientific domains on the basis of a set of patents.
- Trend detection/emerging field detection: Concepts and topics extracted from a set of patents or publications can be identified (first time occurrence) and their evolution over time can be assessed.
- Science-technology spill-overs: Concepts and topics extracted from a set of patent documents can be compared to concepts and topics extracted from a set of scientific publications and vice versa, so that spill-overs between technology and science can be assessed in a fine-grained manner.
- Characterizing the patent and/or publication portfolios of firms, universities, and research institutes in terms of contents and topics.

38.2 Practical Text Mining Procedure

38.2.1 Vector Space Model

Text mining implies a mathematical representation of textual data in such a way that algorithms and data mining techniques can be used. The vector space model (VSM) is a common algebraic representation of text documents. This so-called bag-of-words approach implies that the number of occurrences of each word in a text is counted [38.32–34]. The vector space of a collection of texts is constructed by representing each document as a vector containing the frequencies of the words or terms encountered in that document. Together, these document vectors comprise a term-by-document matrix representing the full text collection. Relatedness of documents can be derived from these vectors, e. g., by calculating the Euclidian distance between two vectors or the angle between document vectors by means of a cosine measure.

38.2.2 Indexing

The encoding of documents into vectors is called indexing. During indexing, a global vocabulary (thesaurus) is built up, assigning a unique identifier to each word encountered in the entire document collection. With this global vocabulary, a vector is constructed for each document comprising as many elements as the total number of unique words in the global vocabulary. For words appearing in the document at hand, the value of the respective elements is equal to the number of occurrences of that word in the document. For words that do not appear in the document, the respective elements are assigned a zero value. Thus, each document is represented by a vector comprising the raw frequencies of occurrences in a (high-dimensional) vector space of terms. As each document uses only a small subset of words to describe its content, the resulting matrix is extremely sparse (typically, more than 99.99% of the vector elements are zero).

38.2.3 Pre-Processing: Feature Selection and Extraction to Deal with Language Issues

Although this numerical representation of the text collection using raw term frequencies within documents enables the application of traditional data mining techniques, such as distance calculation, clustering, and classification (with term vectors as variables and document vectors as observations), results tend to be biased because of typical language issues. First, there are morphological problems, since words can appear in different word forms (e. g., singular versus plural, adjective versus noun versus verb). Second, there is the homonymy, polysemy, and synonymy problem and the hyponymy/hypernymy problem: Words can have different meanings, the same meaning or concept can be expressed by different words, and words can have semantic associations in a hierarchical relation (e. g., animal versus mammal versus cat and dog). Last, not all words are of equal importance in deriving the meaning of phrases, but words with less significance appear with very high-frequency—Zipf's law: The frequency of any word in a text corpus is inversely proportional to its rank in the frequency table [38.21].

As a result, deriving similarity on the original document vectors containing the raw term frequencies of occurring words/terms—and hence, all data mining techniques based on similarity such as clustering and classification—can be biased; two numerically similar document vectors can point to two documents with a completely different meaning because identical words can be used to convey different meanings and because many frequently used words belie the intended meaning. Likewise, two numerically dissimilar document vectors can point to two related documents because meanings and concepts can be expressed through different words, and words can appear in different forms.

To deal with these issues, it may be pertinent not to use the original terms but to engage in prior transformations. Transforming the original data elements into more appropriate features is a common approach in data mining (feature selection and extraction). In practice, for text mining, pre-processing actions such as stemming, stop-word removal, collocations, synonym lists, domain vocabulary, part-of-speech taggers, chi-square tests and information gain, and weighting schemes (e. g., TF-IDF—term-frequency × inverted document frequency—and log-entropy) are commonly used to improve the indexing process and achieve a better grasp of the context of the documents.

Stop-Word Removal and High-Frequency Term Removal

All common words that do not contribute to the distinctive meaning and context of documents might be removed before indexing. Examples include words like and, or, the, a, if, and so forth. Commonly used word lists are available, containing a large set of so-called stop words (e. g., the SMART list of Buckley and Salton, Cornell University). Moreover, as very common words usually convey little information on the meaning of phrases, high-frequency words might be removed as well, by using the distribution as described in Zipf's law.

Stemming and Lemmatization

Instead of indexing words as they appear in the documents, linguistic stems or roots can be used for indexing. The basic idea is to reduce the number of words by introducing a common denominator, called a stem or root, for words that share a common meaning (e. g., am/are/is/be, dog/dogs, or work/workers/working).

The idea behind stemming and lemmatization is to improve the ability to detect similarity regardless of the use of word variants (reduce the number of synonyms, since multiple terms sharing the same stem or root are mapped onto the same concept), but occasionally stemming will create new homonyms due to stemming errors (for a more in-depth analysis of the performance and advantages/disadvantages of stemming, which are language and corpus dependent, we refer to *Lennon* et al. [38.35], *Harman* [38.36], *Krovetz* [38.37], and *Porter* [38.38]).

Lemmatization is done by using a vocabulary or dictionary and applying morphological analysis to match words or terms with the base word or lemma.

Stemming is a more mechanical approach that does not perform a linguistically correct lemmatization but takes a pragmatic/heuristic approach in stripping suffixes from words to combine word variants with shared meanings (e. g., *produc* for product, production, producing). A well-known and commonly used and effective approach is the Porter stemmer [38.39, 40].

Hapax Removal and Low-Frequency Term Removal

Often, a considerable number of words appear only in one document. Such words are called hapaxes and might be removed from the vocabulary because they are of little value in finding communality between documents. Based on that same logic, sometimes words with very low frequencies are removed entirely. Again, Zipf's law can be used to select low-frequency words. Notice that one removes idiosyncratic elements, which might be of special interest as well.

Part-of-Speech Tagging

Instead of selecting or weighting words using frequency counts to reveal relevancy, the part of speech of a word can be used as indicator of relevance. Part-of-speech tagging can be used to classify all words by their part of speech using their context. Next, only nouns or nouns and verbs can be retained as most relevant words. Multiple automated rule-based and stochastic part-of-speech taggers are available (based on probabilistic models, hidden Markov chains, maximum entropy Markov models, perceptrons, support vector machines, ...).

Weighting

Introducing a weighting scheme based on the occurrence and co-occurrence of terms (raw frequencies) might allow us to better distinguish between terms given the specific set of documents under study (e. g., the word *computer* does not reveal the distinctive nature of a certain contribution within a document set covering only papers on computer science). A commonly used weighting scheme is the TF-IDF (term frequency × inverted document frequency) weighting scheme [38.34], in which the raw term frequencies (local weighting) are multiplied by the inverse document frequency (IDF) for that term (global weighting). This results in upgrading the impact of relatively rare terms when calculating distance measures (for a more extensive explanation of local and global weighting, see [38.41])

$$\mathrm{Idf}_i = \log \frac{n}{\sum_{j=1}^{n} \chi(\mathrm{tf}_{ij})}$$
$$\text{with } \chi(t) = \begin{cases} 1 & \text{if tf} > 0, \\ 0 & \text{if tf} = 0 \end{cases} \tag{38.1}$$

and n the number of documents.

Weighting has a similar effect to stop-word removal, since words commonly used across all documents in the document set will be down-weighted compared to medium or even low-frequency words, which carry the most significant distinctive information [38.42]. On the other hand, TF-IDF weighting attributes might introduce extreme weights to words with very low frequencies. Moreover, TF-IDF will not grasp synonyms; hence, weights of commonly used synonyms will be over-rated, as the weights of the individual (synonym) terms will be higher than the weight of the underlying common concept. Despite these shortcomings, TF-IDF weighting is one of the most popular weighting schemes (for alternatives [38.24]).

An alternative weighting schema combining local and global weighting is log-entropy weighting [38.43], where the log of the raw term frequencies plus 1 (to prevent taking the log of 1) (local weighting) are multiplied by the entropy (global weighting)

$$\mathrm{entropy}_i = 1 + \sum_{j=1}^{n} \frac{p_{ij} \log(p_{ij})}{\log(n)}$$
$$\text{with } p_{ij} = \frac{\mathrm{tf}_{ij}}{\mathrm{gf}_i} \tag{38.2}$$

and gf_i the total number of times that term i appears in the entire collection.

Another alternative when the term frequency is not relevant or misleading is binary weighting, where every term frequency greater than 1 is replaced by 1.

Additional Pre-Processing

Additional, more advanced, pre-processing tasks can be performed to further optimize the indexing process. These include proper name recognition and disambiguation, acronym recognition, compound term and collocation detection, feature selection using application-specific domain vocabulary or ontology, and the use of supervised machine learning techniques such as correlation and information gain/entropy (a more detailed description of these topics can be found in [38.44]).

38.3 Specific Text Mining Models

38.3.1 Language Specific Issues and Models

As mentioned previously, natural language text is noisy for a number of reasons, including inconsistencies, typographical errors, author style, and choice of words. It is further complicated by morphological problems and phenomena such as homonymy, polysemy and synonymy, and hyponymy/hypernymy. A numerical representation of text data such as the vector space model does not grasp these particularities.

Furthermore, morphological problems in securing the proper identification of terms, and the fact that not all terms in a text are of equal importance, can be solved by feature selection and extraction techniques as described in the section on pre-processing. On the other hand, there are more fundamental issues with homonymy, polysemy, and synonymy, and hyponymy/hypernyms that are not solved by the preprocessing techniques described. These issues require specific methods to (try to) understand the meaning of words. Examples of such methods are latent semantic analysis, non-negative matrix factorization, formal concept analysis, probabilistic topic models, and Word2vec.

38.3.2 Latent Semantic Analysis (LSA)

Latent semantic analysis (LSA) was developed in the late 1980s at BellCore/Bell Laboratories by *Landauer* and his team in the Cognitive Science Research Department [38.45]:

> Latent semantic analysis (LSA) is a theory and method for extracting and representing the meaning of words. Meaning is estimated using statistical computations applied to a large corpus of text. The corpus embodies a set of mutual constraints that largely determine the semantic similarity of words and sets of words. These constraints can be solved using linear algebra methods, in particular, singular value decomposition.

LSA is a mathematical and statistical approach, claiming that semantic information can be derived from a word-document co-occurrence matrix and that words and documents can be represented as points in a (high-dimensional) Euclidean space.

LSA builds upon semantic similarity and, hence, uses proximity models such as clustering, factor analysis, and multidimensional scaling (see [38.46], for a survey). Discovering latent proximity structure has previously been explored in relation to automatic document indexing and retrieval, using term, and document clustering [38.32, 47, 48] and factor analysis [38.49–51]. LSA builds further on these factor analysis techniques and constructs a concept-by-document matrix using a low-rank approximation of the term-by-document matrix, combining terms into concepts.

LSA rests on the conceptually simple assumption that the representation of any meaningful passage of text must be composed as a function of the representation of the words it contains. Thus, LSA models a passage as a simple linear equation, and a large corpus of text as a large set of simultaneous equations. The solution is in the form of a set of vectors, one for each word and passage, in a semantic space. Dimensionality reduction is an essential part of this derivation.

Optimal dimension reduction is a common workhorse in the analysis of complex problems in many fields of science and engineering. Over 99.99% of cells in the word-by-paragraph or term-by-document matrix representing the documents in the vector space turn out to be empty. This makes the comparison of word or paragraph meanings quite unpredictable. However, after dimension reduction and reconstruction, every cell will be filled with an estimate that yields a similarity between any paragraph and another, and between any word and another. This dimension reduction is crucial and is what accounts for LSA's advantage over most current methods of information retrieval that rely on matching exact words (or terms). It is also what determines its ability to measure the similarity of two essays that might use totally different words.

In practice, LSA is implemented by using singular value decomposition (SVD). A theorem by *Eckart* and *Young* [38.52] states that the rank-k approximation provided by SVD is the closest rank-k approximation

$$\begin{aligned}\|\mathrm{A}-\mathrm{A}_{k(\mathrm{SVD})}\|_2 &= \min_{\mathrm{rank(B)}\le k}\|\mathrm{A}-\mathrm{B}\|_2\\ &= \sqrt{\sigma_{k+1}^2+\cdots+\sigma_n^2}\,. \end{aligned} \tag{38.3}$$

The actual dimensionality reduction is then realized by truncating the SVD decomposition that decomposes the original term-by-document matrix into orthogonal factors representing both terms and documents

$$\mathrm{A}=\mathrm{U}\Sigma\mathrm{V}^{\mathsf{T}}\,, \tag{38.4}$$

with A the original term-by-document matrix, Σ a diagonal matrix with the square roots of singular values of $\mathrm{A}\,\mathrm{A}^{\mathsf{T}}$ and $\mathrm{A}^{\mathsf{T}}\,\mathrm{A}$ ($\sigma_1^2>\sigma_2^2>\ldots>\sigma_n^2$), and U and V containing orthogonal columns of left and right singular vectors, so that only its k largest singular values and corresponding dimensions of U and V are retained

$$\mathrm{A}=\mathrm{A}^{m\times n}\cong\mathrm{A}_k^{m\times n}=\mathrm{U}^{m\times k}\Sigma^{k\times k}\mathrm{V}^{k\times n}\,. \tag{38.5}$$

So, SVD is used to truncate the original vector space model to reveal the underlying or latent semantic structure in the pattern of word usage in order to define documents in a collection. This truncation is considered instrumental to deal with typical language issues such as synonymy, since different words expressing the same idea are supposed to be close to each other in the reduced k-dimensional vector space. Thus, instead of working in the original m-dimensional vector space represented by the original term-by-document matrix $\mathrm{A}^{m\times n}$, it is possible to work with the reduced vector space of lower rank, ignoring all but the first k singular values in Σ and all but the first k columns of U and V. This dimension reduction to k dimensions provided by SVD is the closest rank-k approximation available and might be useful to eliminate *noise* by focusing only on the underlying latent structure.

The k dimensions in the reduced space, or concept space, are now no longer mere words or stems, as in the original vector space, but linear combinations of such linguistic terms or stems. Therefore, the basic unit of analysis has become not just a mere word but a word-and-its-context, a concept (hence, the denomination concept space).

As dimensionality reduction turns the original (extremely) sparse matrix into a smaller full matrix, dimensionality reduction has no positive effect on storage size or calculation time; on the contrary, except for very low values of k, dimensionality reduction will increase the storage size as well as increase the processing time for calculations on the obtained reduced matrix. This makes clear that dimensionality reduction is not about computational simplification but rather a fundamental aspect of the method of dealing with language issues and reducing noise (terms in documents that do not contribute to the meaning of the document or parts of the document). As a consequence, the choice of k is not arbitrary but needs to be chosen carefully to truly represent the underlying latent structure of the data. The choice of the number of concepts to be retained is not straightforward. Current literature recommends 100–300 concepts for large datasets [38.53–55]. For some applications, it may be better to use a subset of the first 100–300 dimensions [38.45]. However, the best choice for k may well be database dependent, as *Berry* and *Browne* [38.56] suggest. We will be revisiting this issue explicitly in this chapter, as several validity exercises will include the assessment of dimensionality reduction (Sect. 38.4). In the context of text indexing, LSA is also referred to as latent semantic indexing (LSI).

38.3.3 Probabilistic Topic Models

Probabilistic topic models such as latent Dirichlet allocation (LDA) [38.57] help to derive the semantic structure from a set of text documents. LDA presumes text documents are derived from a finite set of topics, and that any topic can generate a set of words with a certain probability. So, documents contain multiple topics, and these topics can be represented by multiple words (and any word can be linked to multiple topics with different probabilities). LDA tries to achieve the reverse by deriving the probabilities of every word in the set of text documents being linked to any of the (latent) topics present in the text documents, i. e., the observed distribution of words is used to derive the unobserved distribution of topics. As a result, every word is given a list of probabilities of being linked to any of the topics, and every document is given a list of probabilities of being linked to any of the topics (by combining the probabilities of the words present in the document).

The advantage of this approach is that the individual word is not taken as a feature, but the context of words is taken into account in dealing with synonymy and homonymy/polysemy problems, i. e., the context of words is taken into account to derive the meaning of words. The additional strength of LDA is that words—and, hence, documents—can be linked to multiple topics to effectively deal with homonymy/polysemy issues. Other statistical techniques that try to derive the meaning of words, such as LSA and non-negative matrix factorization, also take the context of words into account but do not address homonymy/polysemy issues. These techniques seek to derive the meaning of words from the context, but they only derive one meaning for any single word. In contrast, the design of LDA is such that multiple meanings can be derived for the same word.

38.3.4 Similarity or Distance Calculation

The similarity measure typically used in information retrieval applications [38.56] is the cosine similarity measure. It is an expression of the angle between vectors, formulated as an inner product of two vectors, divided by the product of their Euclidean norms.

If vectors are normalized beforehand, this formula reverts to the simple inner product. Since, in the original vector space, all vector elements are positive (a word will appear zero times or more in a document), the results are values between 1 (for similar vectors, i. e., pointing in the same direction) and 0 (for entirely unrelated vectors), even after the application of a weighting scheme such as TF-IDF. This yields distances between 0 and 1. However, it no longer holds for vectors in the reduced concept space after SVD (LSA), since vector elements may be negative, which results in a concept-by-document space $V^{k\times n}$ that is no longer positive semi-definite with distances between 0 and (theoretically) 2, although it should be noted that values larger than 1.3 are quite rare in practice. While other similarity measures are possible (e. g., Jaccard, Dice, Euclidean distance—see *Baeza-Yates* and *Ribeiro-Neta* [38.58]), the cosine measure is among the most commonly employed [38.59].

38.4 Document Similarity: Validation Studies

Applying off-the-shelf text mining solutions in science and technology studies is not straightforward. Multiple methods and options are available, but experience with patent and publication data is limited, and more research is needed concerning effectiveness and validity of obtained results. Many options have to be considered during indexing (e. g., stop-word removal, stemming/lemmatization, weighting), modeling and dimensionality reduction (e. g., original document-by-term matrix, LSA, matrix factorization, topic modeling, number of retained concepts/topics), and similarity calculation (e. g., cosine, Jaccard similarity coefficient). Combined, all these options and parameters generate a broad spectrum of possibilities for representing the documents in a vector space and deriving similarity metrics. Although some generally accepted practices exist, there is still a lack of clarity about which options yield better results and under what circumstances.

To obtain further insights into the difference and relevance of options and parameters, we compared the outcome and accuracy of multiple options and parameters in two case studies, a small-scale study on matching patent and publication portfolios of academic inventors and a large-scale study on matching patents and publication for the purpose of identifying patent–publication pairs.

38.4.1 Case Study 1. Comparative Study of Similarity Metrics: Small-Scale Matching of Patents and Publications Within Portfolios of Academic Inventors

In this first case study, we endeavoured to match patent and publications within the portfolios of academic inventors (for a more extensive account, see *Magerman* et al. [38.10]).

For six academic inventors from KU Leuven—four from the Faculty of Medicine and two from the Faculty of Engineering—all WIPO (World Intellectual Property Organization), EPO (European Patent Office), and USPTO (United States Patent and Trademark Office) patents were extracted, where they appeared as inventors. After deduplication of the patent families and removal of patents without abstracts, 30 patents, ranging from 2 to 12 patents per academic inventor, were withheld. Next, we collected all publications of these professors appearing in the Web of Science. This resulted in 437 publications, ranging from 33 to 106 publications per professor (only publications with an abstract were retained). Altogether, the dataset consists of 467 documents.

Similarity Measures

For every patent of a given academic inventor, similarities with all publications of the same inventor are derived based on patent and publication titles and abstracts.

A variety of similarity measures were derived based on following options:

- Stop-word removal: SMART stop-word list plus removal of most frequent words after stemming (*method*, *present*, *result*, *study*, and *type*)
- Stemming/lemmatization: Porter stemmer
- Weighting: No weighting (raw term frequencies) and TF-IDF weighting
- Modeling/dimensionality reduction: No dimensionality reduction (straight cosine on the original document-by-term matrices after weighting) and LSA with $k = 5, 10, 20, 30, 100, 300$ (after weighting)
- Similarity metric: Cosine distance.

Dimensionality reduction is an essential part of the LSA method; it truncates the vector space model to reveal the latent semantic structure in the document collection, mapping terms on latent concepts by combining terms in linear relationships. SVD is applied to obtain a rank-k approximation of the original matrix. Dimensionality reduction is supposed to remove the noise due to polysemy and synonymy present in text documents, but the level of reduction, or the best selection of the rank-k of the truncated document-by-term matrix, remains an open question. Empirical testing suggests that the optimal choice for the number of dimensions ranges between 100 and 300 for large datasets. For small datasets, low values of k (below 10) seem to work equally well [38.3, 60].

In general, a set of documents is indexed and weighted as a whole, and LSA is performed on the index of all documents. We are only interested in relations within the set of patents and publications of the respective academic inventors. As such, there are two options to perform weighting and LSA: perform weighting and LSA on the unified vocabulary of all six academic inventors, or perform weighting and LSA on the case-specific vocabulary of each academic inventor separately. The individual or case-specific approach holds the promise that weighting and LSA are tailored for each professor individually. At the same time, the individual document sets become small; only low k-values can be used for the SVD rank-k approximation.

We include both approaches in our analysis. For the case-specific vocabulary approach, the highest rank-k approximation depends on the size of the smallest document set of all academic inventors, which is 66 (a professor with 2 patents and 33 scientific publications). We opted to include rank-k approximations of 30, 20, 10, and 5 (as previous research on small document sets suggests that even very low values of k might yield relevant results [38.3, 60]). For the global unified vocabulary approach, the highest rank-k approximation possible is 467 (the total number of documents of all academic inventors). We opted to include rank-

k approximations of 300 and 100 and also included 30, 20, 10, and 5 for comparison with the case-specific approach. In the interest of simplicity, we denote the different rank-k approximations by SVD followed by the rank-k approximation (e. g., SVD 30 means we applied LSI using a rank-30 SVD approximation).

Hence, we consider the following options: global unified document indexing (index=U (unified)) and individual case document indexing (index=C (case)); no weighting (weighting=NO) and TF-IDF weighting (weighting=TI); no SVD reduction and reduction to 5, 10, 20, 30, 100, and 300 concepts (the latter two only for the global unified document indexing). Table 38.1 contains an overview of the measures considered.

SVD 0 denotes that no SVD/LSA has been performed, while SVD 30 implies an SVD rank 30 approximation. There are fewer measures pertaining to local case document indexing as it is not feasible to apply higher rank SVD solutions (beyond 300) given the limited size of the datasets implied. It should be noted here that, while Table 38.1 lists 24 combinations, only 23 distinctive measures are implied. When neither weighting nor LSA are applied, global unified document indexing and individual case document indexing yield similar distance measures.

As we are only interested in the relation between patents and publications within the separate document sets of the respective academic inventors, we only calculate the distance between the patents of an academic inventor and all publications of the same academic inventors, and not the distance between patents and publications of different academic inventors. In practice, this means that we have 23 calculations for 2345 different patent–publication pairs.

Table 38.1 Case study 1. Overview of measures

Measure	Index	Weighting	SVD
1	U	NO	0
2	U	NO	5
3	U	NO	10
4	U	NO	20
5	U	NO	30
6	U	NO	100
7	U	NO	300
8	U	TI	0
9	U	TI	5
10	U	TI	10
11	U	TI	20
12	U	TI	30
13	U	TI	100
14	U	TI	300
15	C	NO	0
16	C	NO	5
17	C	NO	10
18	C	NO	20
19	C	NO	30
20	C	TI	0
21	C	TI	5
22	C	TI	10
23	C	TI	20
24	C	TI	30

Comparison of Similarity Values Obtained

For every similarity metric (i. e., for every combination of indexing and similarity calculation options), we compiled descriptive statistics on the distribution of the distance values obtained (distance between every patent and every publication for the six academic inventors—note that this distance becomes zero when two documents are similar, so distance$_{ij}$ is calculated as $1 - \text{cosine}_{ij}$).

Figure 38.1 reveals that one group of measures displays a highly skewed distribution (measures with no or high values of SVD), while other measures are far less skewed (measures with low values of SVD). Figure 38.1 reflects distribution examples for a number of representative measures obtained.

In Fig. 38.1, M1 is the measure with neither weighting nor SVD (LSA), M3 is a measure with no weighting and low SVD (SVD 10) performed on the global document set, M13 is a measure with weighting and medium SVD (SVD 100) performed on the global document set, and M24 is a measure with weighting and medium SVD (SVD 30) performed on the case document set. The lines represent the number of patent–publication and patent–patent pairs with distances within the range indicated on the x-axis (distance buckets of 0.1 expressed as distance—0 is total similarity, 1 is total dissimilarity). The measure with low SVD (M3) is very distinct from the other measures and has a counter-intuitive shape, since one does not expect so many close pairs (more than 1500 pairs have a distance of 0, so are classified as identical). It seems that low SVD maps too many unrelated terms to a small number of con-

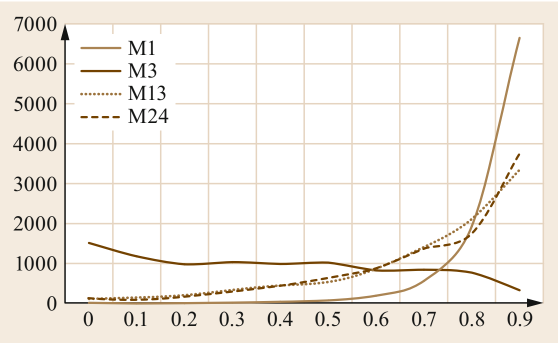


Fig. 38.1 Case study 1. Distribution examples

cepts, artificially creating close pairs. In a following step, we compared the calculated similarity measures with similarity ratings obtained from independent ratings.

Assessment

In order to compare and assess the accuracy of the different measures, all implied patent–paper pairs were rated independently by two researchers. Overall, five researchers—all active and experienced in the field of science and technology studies—were involved in this exercise for all six academic inventors.

Each researcher rated the relatedness between patent documents, on the one hand, and publications, on the other. Three categories were used, ranging from highly related to unrelated, with somewhat related in the middle. In a subsequent step, the scores of each pair were compared, and Kappa scores—indicating between-subject consistency—were calculated. If two assessments differed greatly (highly related versus unrelated), both assessors reviewed and discussed their assessments. After this iteration, Kappa scores were obtained ranging from 0.62 to 0.90, signaling satisfactory and even excellent levels of consistency (the average for the six academic inventors was 0.83).

These assessments figured as the independent variable in an ANOVA (analysis of variance) analysis to assess the relation between the expert assessment, on the one hand, and the relatedness as obtained by the calculated measures, on the other. For the six professors in our study, 16 patents (all patents of four academic inventors and a selection of patents of the remaining two academic inventors—3 out of 9 and 4 out of 12 patents, respectively) were assessed independently in terms of relatedness. Table 38.2 provides an overview of the average R^2 obtained for the measures under study. The higher the observed R^2, the more the similarity measures concur with the independent expert assessments.

Table 38.2 reveals considerable differences between different conditions. If we work with a unified thesaurus, measures that concur most with expert ratings imply the application of TF-IDF weighting with either high SVD values ($n = 300$) or no SVD/LSA at all. Similarly, when opting for a case-based thesaurus, measures that combine weighting, either with no SVD or high levels of SVD ($n = 30$) perform best. Differences with fewer performing combinations are highly significant ($p < 0.0001$). Better performing measures share the characteristic that they are relatively modest in terms of information reduction. Applying no SVD/LSA by definition implies refraining from reducing the initial word space; applying SVD with a relatively large number of dimensions respects the potential richness of the underlying information.

Table 38.2 Case study 1. Accuracy levels obtained for different measures under study

Index	Weighting	SVD	Mean	Std. deviation	N
Case	NO	0	0.401	0.293	16
		5	0.247	0.257	16
		10	0.321	0.254	16
		20	0.362	0.274	16
		30	0.379	0.270	16
	TF-IDF	0	0.459 (3)	0.288	16
		5	0.191	0.203	16
		10	0.356	0.265	16
		20	0.409	0.277	16
		30	0.413 (4)	0.295	16
Unified	NO	0	0.401	0.293	16
		5	0.106	0.135	16
		10	0.195	0.273	16
		20	0.242	0.280	16
		30	0.285	0.324	16
		100	0.341	0.314	16
		300	0.386	0.286	16
	TF-IDF	0	0.489 (1)	0.301	16
		5	0.133	0.185	16
		10	0.202	0.263	16
		20	0.251	0.296	16
		30	0.314	0.335	16
		100	0.340	0.324	16
		300	0.482 (2)	0.285	16

TF-IDF weighing also has a positive impact, albeit smaller than the application of SVD. The positive impact of weighting can be understood as giving emphasis to the distinct elements of documents.

While the observations related to weighting may come as no surprise, the results on SVD are more surprising. As Table 38.2 reveals, SVD performs worst under all circumstances, especially with a limited number of dimensions. The higher the number of dimensions retained, the more the scores approximate the scores with no SVD/LSA applied, but no level of SVD reduction yields higher scores. Given the premises of LSA, we expected better scores for at least some levels of SVD dimensionality reduction. We should bear in mind that the purpose of LSA is to secure better similarities; that means that SVD/LSA is expected to yield better results than a cosine applied to the original document-by-term matrix, at least for a range of k-values.

While the reduction in overall R^2 in Table 38.2 already illustrates how validity fluctuates, scrutinization of specific pairs really reveals the impact of parameter choices. Table 38.3 provides a detailed insight with respect to the distances obtained under different conditions for a patent–paper pair that is highly related and highly unrelated.

Table 38.3 Case study 1. Impact of specific text mining choices on distance measure obtained

Seed patent	Gluten biopolymers
Publication 1 (close to seed patent)	Designing new materials from wheat protein
Publication 2 (far from seed patent)	In-situ polymerization of thermoplastic composites based on cyclic oligomers

Options taken to arrive at similarity measures			**Measures obtained**		**Assessment**
Index	**Weighting**	**SVD**	**PUB 1 (highly related)**	**PUB 2 (unrelated)**	
Unified	NO	5	0.015	0.009	Misleading
Unified	TF-IDF	300	0.102	0.908	Accurate
Case	NO	5	0.051	0.036	Misleading
Case	TF-IDF	30	0.030	0.967	Accurate

It should be noted here that distance is calculated; hence, low values indicate similarity—with a zero value indicating complete similarity; values close to 1 signal no relatedness at all. As Table 38.3 clarifies, applying an SVD solution to a limited number of dimensions ($n = 5$) results in similarity measures that suggest that publication 2 is more related to the patent document than publication 1, whereas, in fact, the opposite holds. This phenomenon manifests itself both when using a unified and a case-based thesaurus. This example illustrates how a strong reduction in underlying information may result in vector spaces that—when used to calculate distances between objects—yield distance measures of a misleading nature. At the same time, the other two examples included in Table 38.3 (unified thesaurus, SVD 300 and case-based thesaurus, SVD 30) strongly demonstrate the feasibility of applying text mining algorithms to detect similarity, even in the case of document sets stemming from different activity realms (patents and publications). Overall, these observations suggest that the choices made, with respect to the set-up of a vector space model and how to proceed when calculating similarity measures, considerably affect the outcomes obtained.

Conclusions

Our findings reveal that different options and methods coincide with considerable differences in terms of accuracy. While several combinations allow us to arrive at practical solutions, certain combinations display low levels of accuracy and even result in misleading similarity measures. For relatively small datasets, options that respect the potential richness of the underlying data yield better results: Either one opts for no SVD or SVD with a relatively high number of dimensions. In addition, weighting has a beneficial impact under these conditions. For a collection of smaller datasets, a global unified indexing and weighting approach does not yield worse results than an individual, case-based indexing and weighting approach. This is an interesting finding because a global unified indexing approach is far easier in practice. LSA seems not to redeem its promise to deal with synonymy and polysemy problems in our setting; all measures involving SVD perform worse than those where SVD is not applied. We suspect this has to do with the low number of documents in the sample, especially given our case-based indexing and SVD approach.

38.4.2 Case Study 2. Comparative Study of Similarity Metrics: Large-Scale Matching of Patent and Publications

In this second case study, we compare more options in a large-scale matching of patents and publications for the purpose of identifying patent–publication pairs—i. e., scientific publications from which the contents (methodology, findings, discovery/invention) are part of a patent publication, in the field of biotechnology—since it is known to be a science-intensive field with substantial science-technology interactions [38.11, 61].

On the patent side, the OECD definition of biotechnology is used to identify biotechnology patents [38.62], defining 30 international patent classification subclasses/groups related to biotechnology. We use PATSTAT (EPO Worldwide Patent Statistical Database) to retrieve all EPO and USPTO granted patents with application and grant year between 1991 and 2008 according to the 30 defined IPC-subclasses/groups (International Patent Classification) related to biotechnology. Since text mining techniques are applied to further identify patent–publication pairs, only patents with titles and a minimum abstract length of 250 characters are withheld, resulting in a final patent data set of 7254 EPO and 80 994 USPTO biotechnology patents; hence, 88 248 patents in total (PATSTAT, edition October 2009).

On the publication side, we select biotechnology publications (articles, letters, notes, reviews) from the Clarivate Analytics (formerly Thomson Reuters ISI) Web of Science database based on the Web of Science subject classification, for the same time period 1991–2008 (volume year). 243 361 publications are extracted from the subject category *biotechnology and applied microbiology*.

To ensure that all potentially related scientific publications are included in the data set, we extend this core publication set with publications from nine related subject categories:

1. *Biochemical research methods*
2. *Biochemistry and molecular biology*
3. *Biophysics*
4. *Plant sciences*
5. *Cell biology*
6. *Developmental biology*
7. *Food sciences and technology*
8. *Genetics and heredity*
9. *Materials for microbiology*.

This results in more than 1.75 million additional publications for the period 1991–2008—a considerable computational challenge for the text mining method to identify patent–publication pairs. In order to lower the number of publications for ease of calculation without losing too many relevant documents, we retain only those publications from this extended set that cite or are cited by at least one publication from our core set, sizing down the additional publication set to 683 674 publications. Finally, we add all multidisciplinary publications from *Nature*, *Science*, and *Proceedings of the National Academy of Sciences of the United States of America*, resulting in 97 970 additional publications. Again, we retain only publication documents with titles and a minimum abstract length of 250 characters, resulting in a final publication set of 948 432 biotechnology-related publications.

Similarity Measures

For biotechnology patents, similarities with all biotechnology publications are derived from patent and publication titles and abstracts.

A variety of similarity measures were derived from the following options:

- Stop-word removal: SMART stop-word list
- Stemming/lemmatization: Porter stemmer
- Weighting: No weighting (raw term frequencies), binary weighting (0 for no occurrence, 1 for one or more occurrences), IDF weighting, and TF-IDF weighting
- Modeling/dimensionality reduction: No dimensionality reduction (straight cosine on the original document-by-term matrices after weighting) and LSA with $k = 5$, 10, 25, 50, 100, 200, 300, 500, 1000 (after weighting)
- Similarity metric: Cosine similarity and three variants of the Jaccard similarity coefficient: the Sørensen–Dice coefficient, the symmetric variant of the Tversky index with $\alpha = 0$ and $\beta = 1$, and the symmetric variant of the Tversky index with $\alpha = 1$ and $\beta = 0$.

To summarize, we compare 40 measures based on LSA by combining 4 levels of term weighting with 9 levels of dimensionality reduction by SVD and no dimensionality reduction at all. For each of these 40 measures, we use the cosine metric to arrive at a similarity value.

We also include three measures based on counting the number of terms that the patent and publication documents have in common. For these measures, we use stop-word removal and stemming, but only take the distinct terms of every document into account (equal to binary weighting for LSA and cosine). To arrive at a metric similarity measure between 0 and 1, three ways of normalization are used:

1. Divide the number of common terms by the minimum of the number of terms of the patent document, on the one hand, and the number of terms of the publication document, on the other (*common terms MIN*).
2. Divide the number of common terms by the maximum of the number of terms of the patent document and the publication document (*common terms MAX*).
3. Divide the number of common terms by the average of the number of terms of both documents (*common terms AVG*).

The second option is more restrictive compared to the first option and only attributes high similarities if both documents are almost identical (intersection of both documents equal to the union of both documents: $A + B \approx A \cap B$). The first option also attributes high similarity if one document is a subset of another document, even if the latter document contains far more information (the intersection of both documents is equal to one of the documents, but the potentially large complement of that one document is neglected: $A + B \neq A \cap B$ but $A \approx A \cap B$). Hence, the first option will yield higher similarity values for the same document combinations than the second option, and the third option will be somewhere in between. Bear in mind that these three metrics are a generalization of the Jaccard similarity coefficient: the symmetric version of the Tversky index with $\alpha = 0$ and $\beta = 1$, and symmetric version of the Tversky index with $\alpha = 1$ and $\beta = 0$, and the Tversky index with $\alpha = 0.5$ and $\beta = 0.5$ (also known as the Sørensen–Dice coefficient). The distances between all seed patents and all target publications are calculated using these different distance-measure variants.

We deliberately decided not to apply more pre-processing tasks, such as compound term and collocation detection, because we wanted to keep the processing simple and automated. These more advanced pre-processing tasks almost always imply greater human involvement and manual attention. In this setting, we wanted to see if a simple automatic approach would work.

Comparison of Similarity Values Obtained: Aggregative Results and Validation with Control Sets

In a first step, we compare the distribution of the similarity values obtained. For each measure, we take the closest publication and the corresponding similarity value for each biotechnology patent; hence, 88 248 similarity values for each measure. Based on these values, we create normalized histograms displaying the proportion of biotechnology patents having a closest biotechnology publication in a given similarity interval (one histogram for every similarity measure).

In order to assess the validity of the measures, we create three control sets for validation containing patents not related to biotechnology:

1. Agriculture
2. Automotive
3. Materials.

For each of these sets, we randomly select 2500 EPO and 7500 USPTO granted patent documents from the same time period based on IPC-codes (respectively IPC class A01 for agriculture, B60 and B62 for automotive, and IPC subclass G01N, G01R, and HO1L for materials). We retain only documents with titles and a minimum abstract length of 250 characters, resulting in a control set of 29 952 patents related to agriculture, automotive, and materials. For these control groups, we again compile normalized histograms displaying the proportion of agriculture, automotive, and material patents having a closest biotechnology publication in a given similarity interval for every similarity measure. Thus, for any given measure, we can combine the histograms of the biotech patents and the patents related to agriculture, automotive, and materials, since we used the relative share of patents that have a closest publication within a given similarity interval and not the absolute number of patents.

Figure 38.2 shows such a combined histogram for the similarity measure using TF-IDF weighting and SVD of rank-300, a commonly used measure. The y-axis contains the proportion of patents that have a closest publication with similarity given by the x-axis (with intervals of 0.05). Five histograms are combined:

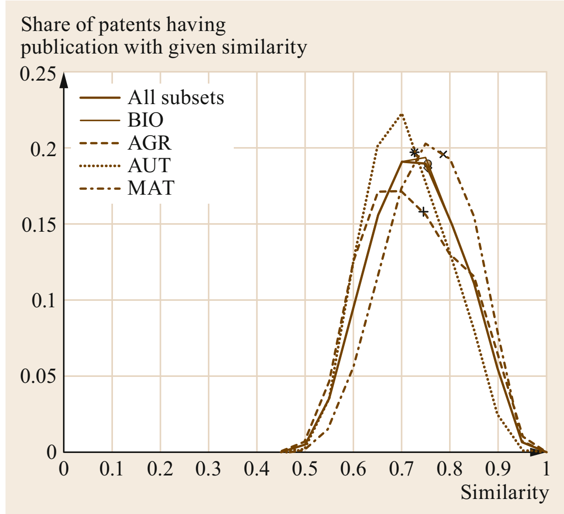


Fig. 38.2 Case study 2. Distribution *patent–closest publication* similarity values according to TF-IDF SVD-300. Frequency count; *markers*=median

one for the biotechnology patents (solid thin line), one for every control group—agriculture (AGR), automotive (AUT), and materials (MAT) (non-solid lines), and one for all patents together—biotechnology patents and all patents from the three control sets (thick solid line).

The distribution of the group of biotechnology patents (solid thin line) falls more or less together with the distribution of all patents (same median value). The relatively high similarity scores are striking. The median similarity for all patents (solid thick line) is 0.76, or 50% of all patents have a scientific publication with similarity above 0.76. These high average similarity values are suspicious, although this could also be considered as a calibration issue. More problematic is the distribution of the similarity measures related to materials patents (dot–dash line). We expect similarity distributions of the sets of control patents to be to the left of the similarity distribution of biotechnology patents. Indeed, those control set patents are, based on the field classification, expected to be less related to biotechnology publications (than biotechnology patents). However, the opposite is suggested in Fig. 38.2. On average, patents belonging to the field of materials are portrayed as more closely related to biotechnology publications than patents belonging to the field of biotechnology. This is very unlikely and suggests that similarity values based on TF-IDF weighting and SVD of rank-300 do not accurately capture the (content) similarity between patent and scientific publication documents.

We observe this phenomenon for all measures based on SVD, and the lower the number of retained dimensions, the worse it becomes (the more distributions shift to the right, the more patents of control groups are more similar to scientific biotechnology publications compared to *biotechnology* patents). Weighting methods have some effect too; distributions based on binary weighting and IDF weighting are shifted more to the left compared to TF-IDF weighting and no weighting at all, regardless of the number of retained dimensions. No weighting at all and binary weighting tend to suffer less from the phenomenon of patents of control groups being more similar to scientific biotechnology publications than biotechnology patents. SVD only seems to yield meaningful similarity values when a high number of dimensions are retained (500 or more) and not in combination with TF-IDF weighting (SVD with 1000 dimensions and TF-IDF weighting still reveals atypical distributions).

Figure 38.3 shows the distribution of the similarity between patents and their closest biotechnology scientific publication according to the similarity measure using TF-IDF weighting without SVD. This distribution makes sense; patents from control groups (agriculture, automotive, materials) are on average less similar to biotechnology patents. Furthermore, there are barely any control set patents that have high similarity with biotechnology patents. The other weighting methods yield similar distributions, although binary and IDF weighting results are slightly more peaked and shifted to the left.

Finally, Fig. 38.4 shows the distribution of the measure based on the number of common terms normalized by the minimum of the term length of both documents (*common terms MIN*). Here, we also observe an expected distribution with control set patents scoring significantly lower similarity values compared to the biotechnology patents. All measures based on the number of common terms yield expected results, although *common terms MIN* yields the highest distinctive power between biotechnology patents and control set patents. Indeed, this measure seems to yield the best distinctive power of all measures under study.

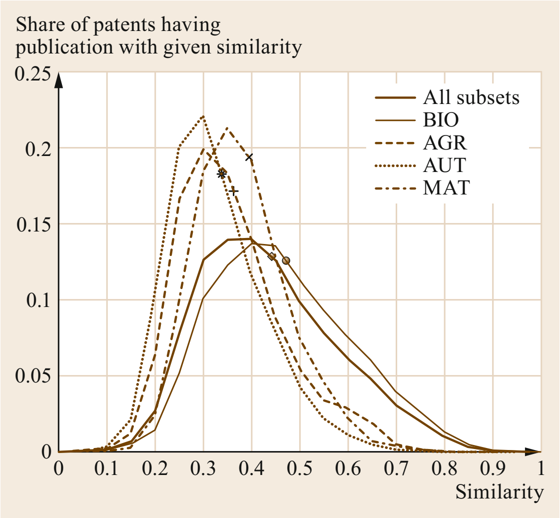


Fig. 38.3 Case study 2. Distribution patent: closest publication similarity values according to TF-IDF without SVD. Frequency count; *markers*=median

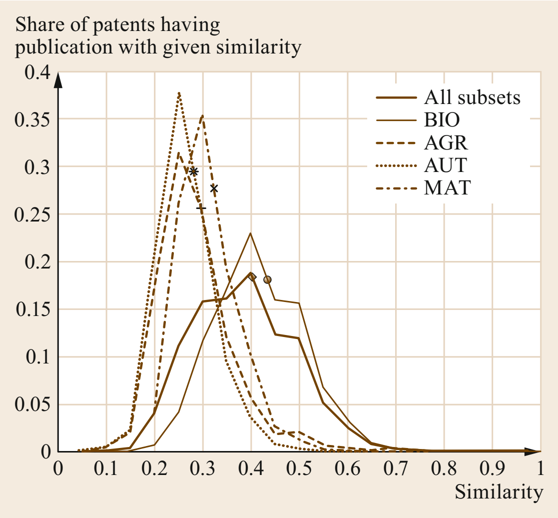


Fig. 38.4 Case study 2. Distribution patent: closest publication similarity values according to number of common terms normalized for minimum term length (common terms MIN). Frequency count; *markers*=median

Preliminary Conclusion

When comparing the distribution of the similarity between patents and their closest scientific biotechnology publication, with the distribution of similarities obtained by comparison with other fields (agriculture, automotive, materials), questions arise regarding the validity of LSA-based measures when matching patent documents and scientific publications. Not only do LSA-based measures yield remarkably high similarities, they also fail to rate non-biotechnology patents as less similar to biotechnology publications than biotechnology patents. This suggests that these measures do not reflect real similarities present in the document collection. The fewer dimensions that are retained, the more the similarity values obtained seem to deviate from the real relations between the documents. This effect is reinforced when using TF-IDF weighting, a commonly used weighting method. Similarity measures based on the cosine metric without dimensionality reduction seem to perform better, in combination with any of the tested weighting schemas, as well as the measures based on counting the common number of terms. The one normalized by the minimum of the term length of both documents (common terms MIN) seems to yield the best results.

These noteworthy results mean that patent–publication combinations yielding high similarity values deserve a closer look. Table 38.4 contains the similarity scores of a patent–publication combination scoring high on TF-IDF in combination with SVD (ranging from 0.928 to 0.995 according to the number of retained dimensions). The title of the patent is:

> Process and rotary milking parlor for the identification of a milking stall and an animal, in particular a cow, in a rotary milking parlor.

The title of the scientific publication is:

> Growth-behavior of lactobacillus-acidophilus and biochemical characteristics and acceptability of acidophilus milk made from camel milk.

The title and abstract of both documents make clear that they are only (very) marginally related; both are about milk, but the patent is about an apparatus for milking, while the publication is concerned with a comparison of cow milk and camel milk with respect to the characteristics of lactobacillus acidophilus fermentation. The similarity scores obtained contain considerable variation among weighting methods and dimension reduction.

Specifically, TF-IDF in combination with SVD yields high scores; other measures yield lower scores, better reflecting the weak relationship between both documents, although all weighting methods yield high values for high levels of dimensionality reduction (low values of k, right-hand side of the table). In general, binary and IDF weighting yield lower scores compared to no weighting or TF-IDF weighting, although there are some exceptions. The measures based on the number of common terms yield low scores (0.10, 0.07, and 0.08 for *common terms MIN*, *common terms MAX*, and *common terms AVG*, respectively), in line with the real similarity between the two documents.

The non-linear relation between similarity scores and dimensionality reduction should be noted; lower numbers of retained dimensions do not necessarily yield the highest similarity scores. For example, the results for raw, i. e., non-weighted, term vectors are interesting; starting from 0.837 for 1000 dimensions, it increases to 0.905 for 300 dimensions, and then decreases to 0.368 for 50 dimensions, to increase again for lower dimensions. This example confirms that the choice of the right level of dimensionality reduction is far from straightforward and that the weighting method has a considerable impact on the results.

Assessment

To assess the validity of LSA-based measures and to obtain greater insight into the contribution of weighting and dimensionality reduction levels on their performance, we set up a validation exercise at the level of individual patent–publication combinations. We selected 250 patent–publication cases with variation in similarity values amongst measure variants. For those 250 cases, we carried out an independent assessment of experts to rate the similarity on a five-level scale, and we checked the consistency between the expert assessment and each of the 43 measures for the 250 selected validation cases. This allowed us to select the best performing measures.

A group of 9 people with experience in the field rated all validation cases (patent–publication combinations) assessing the extent to which the contents of the patent document and the scientific publication cover the same invention/discovery using a five-level scale: not related at all, somewhat related, related, highly related, and identical. The two independent scores were unified by taking the average of the two scores and rounded to arrive again at a five-level score. To deal with potential disagreement amongst raters, two final scores were retained: a conservative one by rounding the average of the scores down to the nearest integer and an optimistic one by rounding the average up to the nearest integer.

Given the expert assessment of the 250 validation cases, an ANOVA-type analysis can be used to check the consistency between the expert scores (conservative and optimistic) and the calculated similarity values. Table 38.5 contains the results of the generalized linear model (GLM) regression based on 250 patent–publication validation cases for the conservative expert score. This table contains the R^2 value of the model

Table 38.4 Similarity scores for patent US7104218 and publication A1994PC04400005 according to various measures

Weighting method	Dimensions retained (SVD)									
	All	1000	500	300	200	100	50	25	10	5
Raw	0.511	0.837	0.873	0.905	0.754	0.391	0.368	0.608	0.691	0.673
Binary	0.083	0.057	0.025	0.023	0.056	0.087	−0.030	0.492	0.763	0.750
IDF	0.095	0.168	0.162	0.260	0.375	0.403	0.504	0.532	0.698	0.738
TFIDF	0.364	0.928	0.973	0.986	0.991	0.991	0.995	0.980	0.959	0.960

Table 38.5 Consistency between (conservative) expert similarity assessment and similarity measures according to measures (R^2 values of GLM regression based on conservative expert score of 250 validation cases)

Measure		R^2	Measure		R^2
Raw	No SVD	*0.61*	Bin	No SVD	*0.77*
	SVD 1000	0.34		SVD 1000	*0.65*
	SVD 500	0.31		SVD 500	*0.63*
	SVD 300	0.30		SVD 300	*0.58*
	SVD 200	0.31		SVD 200	*0.51*
	SVD 100	0.30		SVD 100	0.45
	SVD 25	0.22		SVD 25	0.38
	SVD 5	0.11		SVD 5	0.20
IDF	No SVD	*0.80*	TF-IDF	No SVD	*0.71*
	SVD 1000	*0.63*		SVD 1000	0.45
	SVD 500	*0.57*		SVD 500	0.34
	SVD 300	*0.54*		SVD 300	0.26
	SVD 200	*0.51*		SVD 200	0.21
	SVD 100	0.49		SVD 100	0.17
	SVD 25	0.46		SVD 25	0.14
	SVD 5	0.21		SVD 5	0.11
Common terms (weighted by min number of terms)					*0.82*
Common terms (weighted by max number of terms)					*0.68*
Common terms (weighted by avg number of terms)					*0.75*

for every measure, with the conservative expert score as a dependent variable and the similarity values of the given measure as an independent variable (values in italics are R^2 values higher than 0.50).

Table 38.5 shows that the application of SVD has a negative impact on the performance of similarity measures. For all weighting methods, dimensionality reduction results in significantly lower R^2 values, i. e., less congruence between the similarity score obtained according to the measure and the similarity score as assessed by the experts; the larger the dimensionality reduction, the lower the R^2 values obtained. This is especially the case when no weighting or TF-IDF weighting is used—this is noteworthy, since the combination of TF-IDF weighting and SVD dimensionality reduction is commonly advanced as relevant for text mining purposes. Binary and IDF-weighting outperforms no weighting and TF-IDF-weighting, whether or not SVD is used. The combination of IDF-weighting without SVD, i. e., a cosine metric based on an IDF-weighted document-by-term matrix, yields the highest R^2 (0.80) of all cosine-based measures. It is also striking that simple metrics based on the number of common terms score very highly, and even more so, the metric based on the number of common terms weighted by the minimum number of terms of both documents (*common terms MIN*) yields the highest R^2 value (0.82).

When the optimistic expert scores are used instead of the conservative ones, results remain the same; despite small changes in R^2 (upward for some measures, downward for others), conclusions about SVD, the weighting method, and the best measures are confirmed.

Similarly, when we convert the five-scale expert scores to two-scale expert scores (identical versus non-identical) to focus on the identification of patent–publication pairs, results remain the same.

Our ANOVA results reveal that the similarity measure, *common terms MIN*, best matches our five-scale expert validation. Of course, it comes as no great surprise that measures based on the number of common terms perform that well; the more terms in common, the more one can expect both documents to be similar. Yet, on the other hand, these simple measures based on the number of common terms can miss relevant matches because they do not deal with language-related issues such as homonymy, polysemy, and synonymy. It is worth noting that, despite this lack of complexity, these measures come closest to the expert assessment of similarity—clearly beating LSA measures that claim to deal with typical language issues. Another noteworthy observation concerns the consistency between *common terms MIN* and the presence or absence of a publication author in the list of patent inventors—a strong indication of whether or not the patent and publication are identical, i. e., share the same contents (methodology, findings, discovery). All patent–publication combinations with a similarity of 0.59 or above, according to this measure, have a publication author listed as patent inventor, and all combinations with a similarity of 0.50 or below do not have a publication author listed as patent inventor (with one exception with a similarity value of 0.16). These observations add to the validity of the results obtained.

Patent–Publication Pair Classification

The objective of this study is not only to identify the best measures, or more specifically the best combination of pre-processing options and distance measures, but also to use those insights to identify patent–publication pairs, i. e., scientific publications from which the contents—methodology, findings, discovery/invention—are closely related to a patent document and vice versa.

The results reported corroborate the use of the measure, *common terms MIN*, to classify patent–publications pairs. However, this measure can be misleading when one document is far shorter than the other document. When almost all terms of the patent abstract are present in the publication abstract, a high similarity is obtained when using the minimum number of terms from both documents as the weighting factor. This approach seems to make sense in general; if the abstract of one document is a subset of the abstract of the other document, they can be regarded as identical. However, when one of the documents is too small, or when the difference in length is too great, using the minimum number of terms as the weighting factor might lead to unreliable results (even for human experts, it remains difficult to assess similarity in these cases). A correction is needed for document combinations with one small and one large document. Instead of adding an absolute criterion based on document size, we examined the impact of an additional relative measure, since we already had one measure available: the measure, *common terms MAX*. So, we combined the primary criterion based on the measure, *common terms MIN*, (e. g., above 0.55) with a secondary criterion based on the measure, *common terms MAX*, to correct for doubt cases. It is clear that setting thresholds on the primary criterion (*common terms MIN*) and the secondary criterion (*common terms MAX*) is a trade-off between false positives and false negatives, or precision and recall. To determine good thresholds for *common terms MAX*, 50 additional cases were validated to cover a broad spectrum of combinations of *common terms MIN* and *common terms MAX*.

Table 38.6 contains precision and recall for different thresholds on the primary and secondary criteria (optimal precision, optimal recall, and balanced precision/recall) based on all 300 cases rated by experts (both for the conservative and the optimistic expert scores). Optimal precision scores can be obtained with a recall of approximately 0.51 (numbers in italics on the optimistic side of the table), optimal recall scores can be obtained with a precision figure of approximately 0.81 (numbers in italics on the conservative side of the table), and balanced precision/recall scores of approximately 0.90 are possible for both precision and recall simultaneously (e. g., *common terms MIN* above 0.55 and *common terms MAX* above 0.30).

However, we have to keep in mind that the precision and recall figures listed in Table 38.6 are not representative for the total population because the similarity values obtained for *common terms MIN* and *common terms MAX* are not equally distributed in the validation sample and the total population (the validation sample includes proportionally more identical document combinations). There are many more patent–publication combinations scoring low on the proposed distance measures in the total population, whereas our initial calculation of precision and recall rates from sample data contained an over-representation of patent–publication combinations scoring high on the respective distance measures. Recall rates in particular suffer from this issue. However, the magnitude of the difference between the precision and recall rates averaged over the validation sample and the real rates based on the distribution in the global population heavily depends on the representativeness of the low-scoring cases in the validation set. More validation cases with lower scores are needed to obtain a more reliable estimate of the real precision and recall.

Table 38.6 Precision and recall for different thresholds on primary criterion and secondary criterion (optimal precision, optimal recall, balanced precision), based on conservative and optimistic expert scores for 300 validated cases

		Conservative		Optimistic	
Primary criterion	Secondary criterion	Precision	Recall	Precision	Recall
0.50	0.10	*0.81*	*0.99*	0.88	0.98
0.50	0.32	0.91	0.92	0.94	0.88
0.50	0.61	0.98	0.55	*1.00*	*0.51*
0.55	0.10	*0.82*	*0.98*	0.88	0.97
0.55	0.30	0.90	0.93	0.93	0.89
0.55	0.61	0.98	0.55	*1.00*	*0.51*
0.60	0.10	0.83	0.95	0.98	0.94
0.60	0.29	0.91	0.92	0.94	0.88
0.60	0.61	0.98	0.55	*1.00*	*0.51*

Conclusions

The most notable finding of this study is the poor performance of SVD-based measures, even when commonly used pre-processing options and levels of dimensionality reduction are deployed (e. g., TF-IDF weighting in combination with SVD with 300–1000 dimensions).

When it comes to the influence of weighting, Table 38.5 shows that weighting methods taking into account term frequencies (no weighting at all and TF-IDF weighting) perform worse compared to weighting methods ignoring term frequencies for all levels of dimensionality reduction. In line with these findings, we also witness better performance for the measures based on the number of common terms, measures that also ignore term frequencies.

Looking at individual cases provides us with a measure of insight into the implications of choosing a particular weighting method. In general, including term frequencies is expected to generate better results, since the number of times a given term appears in one document is an indication of the importance of that term in that particular document. However, for our patent–publication document combinations (mostly of a rather moderate length and with highly technical content), this additional notion of importance derived from term frequencies seems to be of less relevance in assessing the similarity of the documents. Indeed, when looking at multiple document combinations, the human judgment on similarity is far more driven by the kinds of terms in the documents rather than the number of times a particular term appears in a document. This observation explains why weighting methods taking into account term frequencies do not perform better, but it does not explain why they perform so much worse. Again, looking at individual cases reveals some additional insights.

First of all, stemming errors and tokenization and parsing issues sometimes cause artificial inflation of term frequencies, magnifying the impact of the underlying stemming and tokenization errors. An example of a patent–publication combination where the amplification of a stemming error results in misleading similarity values for weighting methods, taking into account term frequency, is a patent document about an incubator with external gas feed and a publication about gibberellin metabolism in suspension-cultured cells of *raphanus-sativus*. Both documents have nothing in common, yet score highly on some measures (and score significantly higher on measures that include term frequencies). Both documents have only two (stemmed) terms in common, *feed* and *ga*. However, the stemmed term *ga* occurs 9 times in the patent document and 29 times in the publication document, resulting in high weights when the term frequency is included. However, the stemmed term *ga* in the patent document is a stemming error derived from *gas*, while the stemmed term *ga* in the publication document is an abbreviation of *gibberellin* and has nothing to do with the stemmed term *ga* in the patent document. For weighting methods not taking term frequency into account, this stemming error counts as just one (albeit wrongly) matching term but, for weighting methods using term frequency, this stemming error is magnified and leads to erroneous results.

Secondly, we observe that words with a particular meaning—and, hence, very relevant in the assessment of similarity—tend to have smaller term frequencies compared to natural language words with a more general meaning. For weighting methods including term frequencies, the weight of these more general natural language words becomes too influential in the derivation of similarity by the cosine metric. Weighting terms by their respective IDF values only partially corrects this problem; TF-IDF performs better than no weighting at all, and IDF performs (slightly) better than binary weighting, but TF-IDF still performs worse compared to binary or IDF weighting.

38.5 Clustering and Topic Modeling Case Studies

Structuring and topic detection of text documents is about finding relevant relationships between documents based on the topics described in the text documents. The standard method for (unsupervised) structuring is clustering. Clustering entails combining documents into sets of related documents based on some underlying similarity criterion. The idea is that documents sharing the same contents or topics will be clustered together, which allows one to discover the topics present in the document collection.

However, clustering text data could well be less effective compared to traditional clustering of numerical data. Vectorizing textual data leads to a high dimensional but extremely sparse data space, making it difficult to differentiate in distances; in a high dimensional but sparse feature space, every pair of elements seems distant. On top of this curse of dimensionality, linguistic issues can also bias distance metrics. As described previously, the linguistic phenomena of synonymy, homonymy/polysemy, and hyponymy/hypernymy result in biased feature/term vectors; partial similarity in feature/term vectors does not necessarily imply partial similarity in the underlying text documents, and different feature/term vectors do not imply differences in content. As a result, distance measures directly derived from a vector space of text documents do not always capture the semantic structure of the underlying text documents; the same idea can be expressed in a variety of words, and identical words can have different meanings.

As cluster techniques rely on some sort of distance metric, it is by no means straightforward for cluster techniques to effectively deal with the semantic issues within text data and produce relevant clusters.

Probabilistic topic models such as LDA help to derive the semantic structure from a set of text documents. By definition, they directly reveal the latent topic structure of a document collection (LDA presumes text documents are derived from a finite set of topics and that any topic generates a set of words with a certain probability, so the result of LDA is a linkage between words and (latent) topics, on the one hand, and documents and (latent) topics, on the other). In this respect, it should be noted that each document can be linked to multiple topics.

We compared both methods in a small case study on structuring a data set that we had used previously (patents from biotechnology, agriculture, automotive, and materials).

38.5.1 Case Study 3. Comparative Study of Topic Detection: Clustering

In this study, we applied clustering on a set of patent documents from four different technological fields:

1. Biotechnology
2. Agriculture
3. Automotive
4. Materials (field selection based on IPC technology classification).

For every field, we randomly selected 10 000 patents (with no overlap, every patent document in the sample is only part of one of four fields).

To validate the stability/generalization of the models, we split the data set into two samples and ran all models twice, once on each half of the dataset. This resulted in two subsamples, each with 20 000 patents, evenly distributed over the four technological fields (5000 biotechnology, 5000 agriculture, 5000 automotive, and 5000 materials). This allowed us to compare results for different data sets.

Again, we began with the titles and abstracts, removed stop words (SMART stop-word list, Salton and Buckley) and applied stemming (Porter stemmer). On this occasion, we compiled a document-by-term matrix in two flavors: one based on the raw term frequencies and one based on TF-IDF weighting.

Figure 38.5 contains the word clouds for each of the four technological fields for the first subsample (after stemming). The words displayed in the word clouds clearly reflect the four technological fields and are quite distinctive in each case. For example, *cell*, *protein*, *sequence*, *peptide*, *DNA*, *acid* for biotechnology; *plant*, *seed*, *flower*, *fish* for agriculture; *vehicle*, *control*, *wheel*, *brake* for automotive; and *layer*, *surface*, *semiconductor*, and *substrat* for materials.

Method

As the method is intended for use on large data sets with millions of patents, we chose k-means as one of the most used non-hierarchical cluster methods. As the Euclidean distance is not appropriate in the case of high-dimensional data, we used spherical k-means clustering. We clustered term vectors derived from patent titles and abstracts and assessed the relevance of the obtained clusters in a quantitative way. As we know, the samples contain patents from four technological fields and we set the number of clusters to four. The genetic algorithm was used for hard clustering, with 50 runs. The clustering was performed twice for each subsample, once for the document-by-term matrix based on TF values (term frequency) and once for the document-by-term matrix based on TF-IDF values.

Validation

The key issue with unsupervised techniques is the validation of the clusters/topics obtained, i. e., whether obtained clusters/topics are relevant. We wish to escape reliance on internal cluster evaluation methods (e. g., homogeneity, silhouette), since they may not be very meaningful in the case of text data (if clustering cannot grasp the semantics of underlying patent data, intrinsic evaluation measures are irrelevant), and since these intrinsic evaluation measures sometimes use the same criteria adopted by the cluster method to allocate objects. Intrinsic cluster evaluation methods can shed light on the density or homogeneity of clusters but not on the relevance of clusters. We anticipate the same problems with internal LDA evaluation methods (log-likelihood, perplexity).

External validation is needed to assess the relevance of the clusters/topics obtained. However, a manual, expert-driven assessment is very time consuming, given the magnitude of the data. Moreover, as multiple

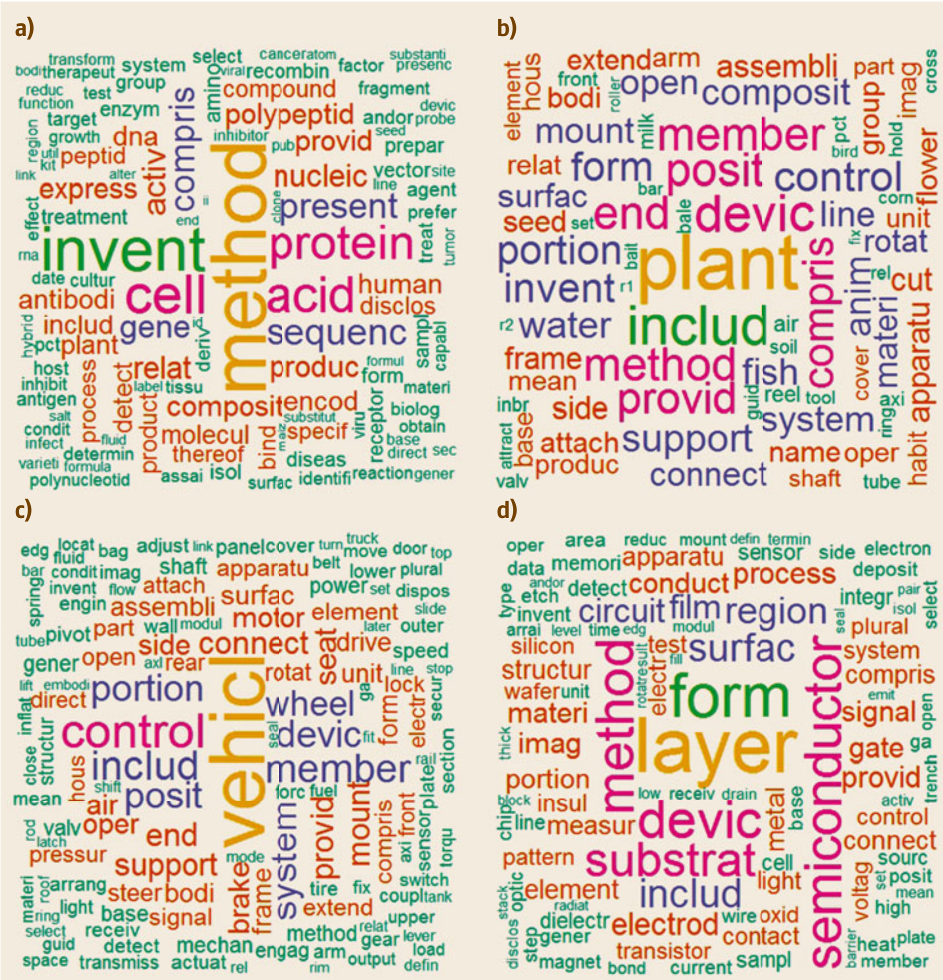


Fig. 38.5a–d Word clouds of the four technological fields for the first subsample: **(a)** biotechnology; **(b)** agriculture; **(c)** automotive; **(d)** materials

parameters/models need to be evaluated, multiple cluster/topic solutions need to be validated. This requires an automated approach in order to quickly asses multiple models.

In our validation setup, we made use of the known (a priori) technology fields of the patents in our data set (biotechnology, agriculture, automotive, and materials). As every patent in our data set is part of one of the four technology fields used to make the patent selection, a cluster solution or topic solution should coincide with the original patent technology classification. By comparing the known patent technology class with the cluster/topic allocation generated, the relevance of the allocations generated can be validated.

Results

Table 38.7 contains the contingency tables for the four cluster solutions. The left-hand side contains the solutions based on TF values, the right-hand side those based on TF-IDF values; the top contains the solutions for the first subsample, the bottom for the second subsample. Each contingency table contains the number of patents per cluster, for each technological field. For example, in the first sample based on TF values, 191 of the 5000 biotechnology patents were classified in cluster 1, 4431 in cluster 2, 28 in cluster 3, and 150 in cluster 4.

The contingency tables show that all four cluster solutions accurately identify biotechnology patents and that, at least, TF-based clustering clearly identifies automotive patents, while agriculture and materials patents are scattered over multiple clusters. However, in a pure unsupervised setup without a priori knowledge of the real patent characteristics, we can only rely on the cluster information. With regard to the technology classification, all cluster solutions reveal rather heterogeneous clusters; for every cluster, patents are scattered all over the technological fields, with the sole exception of materials. Each cluster solution seems capable of generating

Table 38.7 Contingency tables for cluster solutions

Subsample 1, based on TF values

TECH	CLUS1	CLUS2	CLUS3	CLUS4
BIO	191	4431	228	150
AGR	2647	1162	992	199
AUT	4697	111	1	191
MAT	1605	462	1	2932
Total	9140	6166	1222	3472

Subsample 1, based on TF-IDF values

TECH	CLUS1	CLUS2	CLUS3	CLUS4
BIO	70	4562	88	280
AGR	2264	1965	104	667
AUT	3042	58	111	1789
MAT	361	125	2731	1783
Total	5737	6710	3034	4519

Subsample 2, based on TF values

TECH	CLUS1	CLUS2	CLUS3	CLUS4
BIO	159	209	4494	138
AGR	2639	982	1174	205
AUT	4687	0	109	204
MAT	1627	3	564	2806
Total	9112	1194	6341	3353

Subsample 2, based on TF-IDF values

TECH	CLUS1	CLUS2	CLUS3	CLUS4
BIO	294	94	56	4556
AGR	1430	82	2298	1190
AUT	1534	119	3246	101
MAT	1680	2731	373	216
Total	4938	3026	5973	6063

a rather homogeneous cluster for materials, i. e., a cluster with high precision (but, yet, low recall).

Differences between clustering based on term vectors with TF values and clustering based on term vectors with TF-IDF values are not pronounced in this respect. TF values yield two clusters that combine biotechnology and agriculture, one materials cluster, and one leftover cluster. TF-IDF values yield only one biotechnology/agriculture cluster and one cluster that combines agriculture and automotive (and, again, one materials cluster and one leftover cluster). It is striking that using TF-IDF values yields more balanced clusters, i. e., the total number of patents per cluster is more evenly distributed.

Results are remarkably consistent over the two subsamples; both subsamples reveal the same pattern (note that it is not the exact cluster label that is relevant but the distribution over the cluster; in the case of the TF-based solution, cluster 1 and cluster 4 contain the same kind of documents in the two subsamples, while clusters 2 and 3 are switched).

In general, the cluster solutions are not in line with the technology classification. That so many documents from different technological fields are clustered together means other relations have been found. The question is whether these relations are relevant, or whether they are artefacts due to linguistic issues. There are clearly interesting indications that clustering can reveal new insights; equally, it would be unwise to ignore those indications that suggest clustering is inappropriate for (patent) text documents.

However, although cluster solutions are not in line with the technology classification at a first glance, cluster results are clearly not random; all cluster solutions produce one cluster where a very heavy emphasis on biotechnology is combined with a heavy emphasis on agriculture, which makes sense, since agriculture and biotechnology do have some relationship, and one cluster with a mixture of agriculture and automotive, which is a less obvious link. Indeed, they all produce one cluster with a sole emphasis on materials, and finally one leftover cluster. However, although these patterns may make sense, it is not possible to derive a relevant interpretation when no a priori knowledge is available. Indeed, without the known technological classification, it would be impossible to establish which cluster is homogeneous, which cluster links two fields, and which cluster contains leftovers to be reclassified.

38.5.2 Case Study 4. Comparative Study of Topic Detection: Probabilistic Topic Modeling

In this study, we applied probabilistic topic modeling on the exact same set of patents documents of the four above-mentioned technological fields (biotechnology, agriculture, automotive, and materials) to assess probabilistic topic model performance by comparing resulting clusters and topics. Data pre-processing was the same, so the exact same document-by-term matrix was used.

Method

LDA was used to derive topics from the two subsamples independently. Again, we set the number of topics to four and used two methods to estimate probabilities: Gibbs sampling [38.63] and variational expectation maximization (VEM) [38.57], each with five runs. For the Gibbs sampling, we took 1000 burn-in iterations and again 1000 iterations. As LDA requires integer counts, we only used the document-by-term matrix with raw term frequencies.

Results

Table 38.8 contains the contingency tables for the four LDA solutions. The left-hand side contains the solutions based on Gibbs sampling, the right-hand side those based on VEM; the top contains the solutions for the first subsample, the bottom for the second subsample. Each contingency table contains the number of patents per topic, for each technological field. To illustrate this, in the first sample based on Gibbs sampling, 61 of the 5000 biotechnology patents were linked to topic 1, 4575 to topic 2, 247 to topic 3, and 117 to topic 4. It should be borne in mind, however, that, with topic models, documents are linked to multiple topics with different degrees of probability. To construct the contingency tables, we retained only the most likely topic for every document (we will return to this later).

As for the cluster solutions, the contingency tables again reveal that all four LDA solutions clearly identify biotechnology patents, while the other technological classes are scattered over multiple topics (except for the VEM solution in subsample 2, where *materials* is clearly identified instead of *biotechnology*). Again, if we rely only on the topic information, the topics revealed are not consistent with the technology fields. However, once again the same patterns evident in the cluster solutions are disclosed: one topic with a combination of biotechnology and agriculture, one topic with a combination of agriculture and automotive, one topic with materials, and one leftover topic (although, again, the VEM solution on subsample 2 is deviant).

Differences between LDA solutions based on Gibbs sampling or on VEM are more pronounced. In the first subsample, the results for the *biotechnology–agriculture* topic and the *agriculture–automotive* cluster are similar, while characteristics of the *materials only* and the *leftover* topic differ. In the second subsample, patterns differ with a less pronounced link between biotechnology and agriculture. Again, we see a remarkable consistency between the two subsamples with the LDA solutions using Gibbs sampling but not with VEM. It seems that the estimation method does matter, and it would appear that Gibbs sampling yields more stable results than VEM (although we would need further sampling to confirm this).

For our practical purposes—unsupervised extraction of topics—LDA does not seem to do a significantly

Table 38.8 Contingency tables for LDA solutions

Subsample 1, based on Gibbs sampling

TECH	TOP1	TOP2	TOP3	TOP4
BIO	61	4575	247	117
AGR	2129	1944	161	766
AUT	2918	45	182	1855
MAT	355	148	3699	798
Total	5463	6712	4289	3536

Subsample 1, based on VEM

TECH	TOP1	TOP2	TOP3	TOP4
BIO	4575	153	196	76
AGR	1192	892	404	2512
AUT	61	84	1645	3210
MAT	173	2676	1783	368
Total	6001	3805	4028	6166

Subsample 2, based on Gibbs sampling

TECH	TOP1	TOP2	TOP3	TOP4
BIO	176	150	4633	41
AGR	720	199	1901	2180
AUT	1692	230	35	3043
MAT	1414	3138	149	299
Total	4002	3717	6718	5563

Subsample 2, based on VEM

TECH	TOP1	TOP2	TOP3	TOP4
BIO	947	3759	73	221
AGR	1910	194	2689	207
AUT	87	49	4348	516
MAT	163	93	393	4351
Total	3107	4095	7503	5295

better job than clustering; in general, patterns revealed by the topic models are in line with the patterns revealed by the cluster solutions. The fact that topic patterns deviate from the technology classification does not mean that results are irrelevant; the topic distribution obtained can make sense, but it is not possible to make a relevant interpretation when no a priori knowledge is available (which topic is homogeneous, which topic links two fields, and which topic contains leftovers to be reclassified). In a normal unsupervised setup, additional analysis of underlying words and documents is needed to assess relevance.

Figure 38.6 contains the top 20 words for each of the 4 topics in the first subsample based on Gibbs sampling (words with the highest probability of being linked to the topic).

Figure 38.6 confirms that almost all words in topic 3 can be linked to materials, words of topic 4 can be linked to automotive and materials, and words in topic 2 can be linked to biotechnology and agriculture. In topic 1, word links are less pronounced. It seems topic 1 combines agriculture and automotive, because those words can appear in both contexts, while topic 2 combines biotechnology and agriculture, because there is some intersection. In the case of topic 4, it is less clear why material patents are included but are not part of cluster 3. This quick analysis seems to suggest that LDA is able to detect relevant links (the interplay between biotechnology and agriculture) but is not always able to disentangle the context of words (the combination of *agriculture* and *automotive* and, to a lesser extent, *materials* and *automotive*). A more detailed

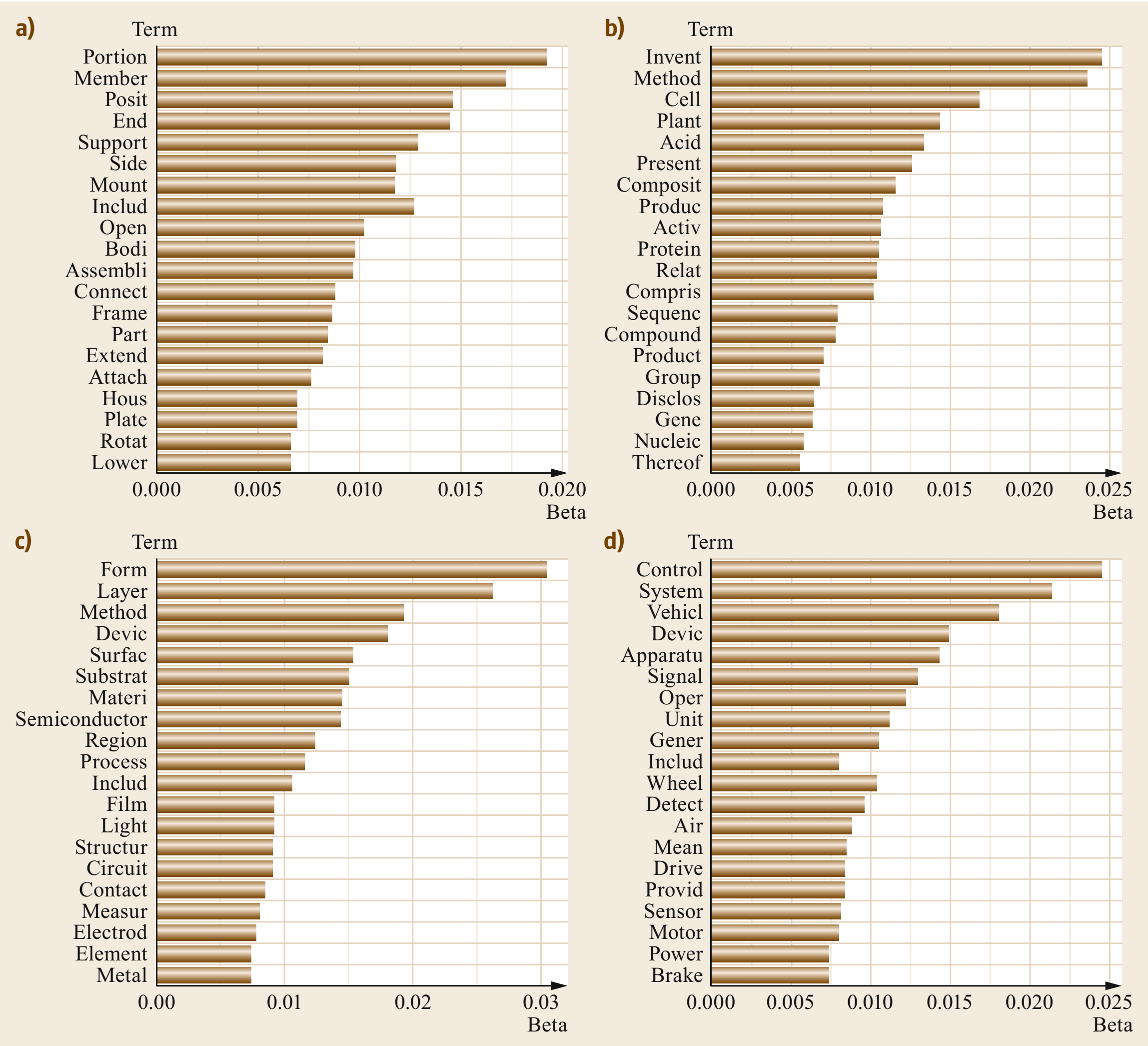


Fig. 38.6a–d Top 20 words for every topic (subsample 1, based on Gibbs sampling). **(a)** Topic 1; **(b)** Topic 2; **(c)** Topic 3; **(d)** Topic 4

analysis of the underlying documents is needed to confirm this.

38.5.3 More Clusters/Topics

More fine-grained clusters/topics might be needed to disentangle technologies; if biotechnology and agriculture are closely related, and agriculture and automotive are somewhat related, a four-topic solution might be inappropriate in revealing our technological patterns. Since more clusters/topics might be needed to reveal subtopics, we designed cluster and topic models with more clusters/topics (e. g., 10, 25). This only partially solved the heterogeneity in found clusters/topics. Some cluster/topics are clearly more homogeneous, but yet a large number of patents are still attributed to heterogeneous clusters/topics.

38.5.4 Retention of Only Single-Topic Documents

Topic models assign all documents to all topics with a given probability. All contingency tables listed previously are based on the primary or main topic of each document, i. e., the topic with the highest probability. The fact that documents are linked to multiple topics can explain the difficulties in finding a good fit between the topic solutions based on the primary topics and the technological classification. To isolate the effect of documents linked to multiple topics, or topics only linked with a very low probability, we analyzed the influence of document selection based on the probability of the topic allocation. We compiled multiple contingency tables where only documents with increasing certainty or reliability of topic allocation are retained, i. e., in compiling the contingency table, we retained only documents with increasing levels of probability of allocation to the topic with the highest probability, combined with increasing levels of the difference in probability between the topic with the highest probability and the next topic. For example, a document with the highest likelihood of being linked to a topic with a probability of 70%, and 8% probability for the second most likely topic, is regarded as a reliable or certain allocation. On the other hand, a document with the highest likelihood of being linked to a topic with a probability of 50%, and then 45% for the second most likely topic, is not regarded as a reliable or certain allocation for that first topic.

Table 38.9 contains an overview of every topic allocation for the first ten documents. Document 1 is an example of a certain or reliable allocation to the primary/main topic: The most likely topic is topic 2 (column T1), with a probability of 0.58 (column P1), while the next likely topic is topic 3 (column T2), with a probability of 0.15 (column P2). Document 3 is an example of an uncertain or unreliable allocation to the primary/main topic: The most likely topic is topic 3, with a probability of 0.33, while the next likely topic is topic 2, with a probability of 0.32. Both topic 2 and topic 3 can, therefore, be regarded as the primary/main topics for this document.

We compiled contingency tables for different degrees of certainty based on the absolute values of the probability of the most likely topic (column P1) and the difference between the probability of the most likely topic and the next likely topic (column D12), but each contingency table resulted in topic allocations that are very close to the one that took all primary topics into account.

38.5.5 Mixed Approach: Cluster Topics

Another approach is to mix both methods. First, derive topics from the documents and then cluster on the retrieved topics instead of the original (weighted) term frequencies. We experimented with a topic model containing 100 topics, followed by clustering with four clusters but, once more, we found results that were remarkably close to the topic model solution incorporating four topics.

Table 38.9 Overview. Every topic allocation for first ten documents

Doc	Topic	Tech	T1	T2	T3	T4	P1	P2	P3	P4	D12	D23
1	2	Bio	2	3	1	4	0.58	0.15	0.14	0.14	0.43	0.01
2	2	Bio	2	4	1	3	0.55	0.17	0.16	0.12	0.38	0.01
3	3	Bio	3	2	4	1	0.33	0.32	0.21	0.13	0.01	0.11
4	2	Bio	2	3	4	1	0.31	0.25	0.23	0.21	0.05	0.03
5	2	Bio	2	4	3	1	0.39	0.23	0.20	0.17	0.16	0.04
6	2	Bio	2	4	1	3	0.56	0.18	0.15	0.13	0.38	0.03
7	2	Bio	2	3	4	1	0.37	0.24	0.21	0.18	0.13	0.03
8	2	Bio	2	3	4	1	0.38	0.25	0.20	0.17	0.13	0.05
9	2	Bio	2	3	4	1	0.39	0.28	0.18	0.15	0.11	0.10
10	3	Bio	3	1	2	4	0.46	0.24	0.15	0.15	0.22	0.09

38.5.6 Conclusions on Clustering and Probabilistic Topic Modeling

In this study, we have sought to investigate the potential of two unsupervised techniques—clustering (spherical *k*-means clustering) and probabilistic topic modeling (LDA)—in order to reveal relations in a large patent collection without a priori knowledge. We did so by using a sample compiled of documents from four different technological fields and by conducting a quantitative analysis on the resemblance between the cluster and topic structure, on the one hand, and the technological fields, on the other.

The results obtained are remarkably stable amongst runs, parameters, and subsamples. All models yield the same pattern in both subsamples, except for LDA models based on VEM. Furthermore, the number of topics retained, the number of iterations for the Gibbs sampling, the distribution of term occurrence over the documents (only retaining terms with distinctive power), and the certainty of allocation of a document to a given topic (topic models assign documents to all topics with a given probability) do not seem to influence results (although further sampling is needed to obtain confirmation).

We do not observe a clear difference between cluster solutions and LDA solutions in retrieving homogeneous and relevant clusters/topics; both methods yield remarkably similar patterns. However, none of the methods clearly reveals the technological fields in the data sample, although both methods reveal patterns that are clearly not random.

It is not because the solutions obtained fail to match the technological fields within the data sample that obtained cluster or topic results lack relevance. The fact that all models yield more or less the same pattern might well indicate that relevant relations beyond the predetermined technological fields are revealed, although it is also conceivable that both methods suffer from the same linguistic issues. A more qualitative study of the data is required to further assess the relevance of obtained clusters/results: Are the documents in the homogeneous materials cluster and the homogeneous materials topic the same? Are the documents within the mixed agriculture/biotechnology cluster and the mixed agriculture/automotive topic assigned to the same agriculture and biotechnology subcluster and subtopic? Can we find a link between the agriculture and biotechnology patents combined in the same cluster/topic when we manually inspect those patent documents? And, equally so, for the less obvious link between *agriculture* and *automotive*? A quick analysis based on the top linked words reveals that biotechnology and agriculture patents are combined because there is a degree of overlap between the two fields, while agriculture and automotive are combined because the same words occur in different contexts.

Unless this more qualitative study is conducted, it is not possible to say whether one method is clearly better than the other, and whether deviations from the technological classification are due to unknown but relevant relations or stem from linguistic issues. The quick analysis based on top linked words suggests that both phenomena occur. For one topic, an interplay between biotechnology and agriculture is discovered while, for another topic, the link between agriculture and automotive seems to be an artefact. Hence, LDA would seem incapable of disentangling all word contexts.

The patterns revealed of both techniques are so similar that we are currently inclined to argue that LDA is not superior in dealing with the linguistic issues that might hamper traditional clustering of document term vectors. For very large data collections, clustering might be considered superior to topic modeling because of the higher computational efficiency of clustering versus topic modeling. On the other hand, topic models could still be relevant, because this technique explicitly assigns documents to multiple topics, which has the appearance of a more natural model for a text document collection. Yet, again, spherical *k*-means also permits fuzzy clustering with documents assigned to multiple clusters. In that respect, it is worth investigating whether topic models or fuzzy clustering can be instrumental in deriving more appropriate distance measures (deriving the similarity of documents based on the similarity in (fuzzy) cluster or topic allocation instead of directly based on the document term vectors).

38.6 Conclusions, Discussion, Limitations, and Directions for Further Research

In this chapter, we outlined the building blocks required to engage in text mining in order to assess the similarity of documents and to structure large text document collections: parsing text documents into terms, pre-processing those terms (stop-word removal, stemming/lemmatization, weighting), compiling a document by term matrix, calculating similarities and, finally, structuring documents into clusters or topics.

All these steps involve a number of choices and, hence, yield a multitude of outcomes. When applying text mining techniques to documents typically used in the STI domain (most notably patents and publications), our findings reveal that these choices have non-trivial consequences. The cases discussed reveal a significant impact on the similarities obtained; the same pair of documents can be classified as either related or unrelated depending on the choices made. Even well-established choices (stemming, TF-IDF weighting) do not always result in valid similarity indicators. Similar conclusions relate to SVD-based measures. LSA is one of the methods intended to derive meaning from the context of words, and it is advocated for its ability to yield better similarity metrics. Our validation studies reveal that the dimensionality reduction involved may result in inferior similarity metrics; the fewer dimensions that are retained, the more the similarity values obtained seem to be misleading (by overestimating the underlying similarity). Furthermore, the promise of probabilistic topic models—another method intended to deal with specific language issues—is not upheld by our observations; the patterns observed do not differ significantly from traditional (spherical) cluster approaches.

Combined, our observations strongly suggest that, in order to harvest the potential of text mining in the field of STI studies, we should refrain from merely adopting off-the-shelf text mining solutions. Investing in validation efforts will remain on the agenda for years to come, e. g., comparing viable approaches based on full text documents (patents/publications/company documents pertaining to new product development/innovation strategy, and so on). While this premise clearly implies additional efforts and resources, we remain convinced that such efforts will pay off. Validated approaches will inform us of a multitude of topics and issues in an unprecedented way. We hope that this chapter inspires colleagues to become involved in unlocking the potential of text mining in our field of research.

References

38.1 D. Hicks, B.R. Martin, J. Irvine: Bibliometric techniques for monitoring performance in technologically oriented research—The case of integrated-optics, R&D Manag. **16**(3), 211–223 (1986)

38.2 S. Hinze, H. Grupp: Mapping of R&D structures in transdisciplinary areas: New biotechnology in food sciences, Scientometrics **37**(2), 313–335 (1996)

38.3 P. Glenisson, W. Glänzel, O. Persson: Combining full-text and bibliometric indicators. A pilot study, Scientometrics **63**(1), 163–180 (2005)

38.4 V. Rabeharisoa: A special mediation between science and technology: When inventors publish scientific articles in fuel cells. In: *Dynamics of Science-Based Innovation*, ed. by H. Grupp (Springer, Berlin, Heidelberg 1992) pp. 45–72

38.5 E.C.M. Noyons, A.F.J. van Raan, H. Grupp, U. Schmoch: Exploring the science and technology interface—Inventor author relations in laser medicine, Res. Policy **23**(4), 443–457 (1994)

38.6 U. Schmoch: The technological output of scientific institutions. In: *Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies of S&T Systems*, ed. by H.F. Moed, W. Glänzel, U. Schmoch (Springer, Dordrecht 2004) pp. 717–731

38.7 W. Glänzel, M. Meyer, B. Schlemmer M. du Plessis, B. Thijs, T. Magerman, K. Debackere, R. Veugelers: Biotechnology: An Analysis of Patents and Publications, Report Steunpunt O&O Statistics, http://www.steunpuntoos.be (2004)

38.8 M. Meyer: Patent citations in a novel field of technology: What can they tell about interactions of emerging communities of science and technology?, Scientometrics **48**(2), 151–178 (2000)

38.9 T. Magerman: *Impact and Consequences of Science-Intensive Patenting: In Search of Anti-Commons Evidence Using Latent Semantic Analysis (LSA) Text Mining Techniques* (KU Leuven, Leuven 2011), unpublished Ph.D. manuscript

38.10 T. Magerman, B. Van Looy, X. Song: Exploring the feasibility and accuracy of Latent Semantic Analysis based text mining techniques to detect similarity between patent documents and scientific publications, Scientometrics **82**(2), 289–306 (2010)

38.11 T. Magerman, B. Van Looy, K. Debackere: Does involvement in patenting jeopardize one's academic footprint? An analysis of patent-paper pairs in biotechnology, Res. Policy **44**(9), 1702–1713 (2015)

38.12 M.A. Hearst: Untangling text data mining. In: *Proc. 37th Annu. Meet. Assoc. Comput. Linguist., College Park, Maryland* (1999) pp. 3–10

38.13 W. Fan, L. Wallace, S. Rich, Z. Zhang: Tapping the power of text mining, Communication ACM **49**(9), 77–82 (2006)
38.14 K.A. Vidhya, G. Aghila: Text mining process, techniques and tools: An overview, Int. J. Inform. Technol. Manag. **2**(2), 613–622 (2010)
38.15 D.R. Swanson: Fish Oil, Raynaud's syndrome, and undiscovered public knowledge, Perspect. Biol. Med. **30**, 7–18 (1986)
38.16 D.R. Swanson: Migraine and magnesium: Eleven neglected connections, Perspect. Biol. Med. **31**, 526–557 (1988)
38.17 D.R. Swanson: Somatomedin C and arginine: Implicit connections between mutually-isolated literatures, Perspect. Biol. Med. **33**, 157–186 (1990)
38.18 N.M. Ramadan, H. Halvorson, A. Vandelinde, S.R. Levine: Low brain magnesium in migraine, Headache **29**(7), 416–419 (1989)
38.19 D.R. Swanson, N.R. Smalheiser: An interactive system for finding complementary literatures: A stimulus to scientific discovery, Artif. Intell. **91**, 183–203 (1997)
38.20 P. Grzybek, E. Kelih: Anton S. Budilovic (1846–1908)—A forerunner of quantitative linguistics in Russia?, Glottometrics **7**, 94–97 (2004)
38.21 G.K. Zipf: *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology* (Addison-Wesley, Cambridge 1949)
38.22 R.E. Wyllys: Measuring scientific prose with rank-frequency ('Zipf') curves: A new use for an old phenomenon, Proc. Am. Soc. Inform. Sci. **12**, 30–31 (1975)
38.23 M. Callon, J.P. Courtial, W.A. Turner, S. Bauin: From translations to problematic networks—An introduction to co-word analysis, Soc. Sci. Inf. (Paris) **22**(2), 191–235 (1983)
38.24 C.D. Manning, H. Schütze: *Foundations of Statistical Natural Language Processing* (MIT Press, Cambridge 2000)
38.25 M.W. Berry (Ed.): *Survey of Text Mining* (Springer, New York 2003)
38.26 E. Leopold, M. May, G. Paaß: Data mining and text mining for science and technology research. In: *Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies of S&T Systems*, ed. by H.F. Moed, W. Glänzel, U. Schmoch (Springer, Dordrecht 2004) pp. 187–213
38.27 A.L. Porter, N.C. Newman: Patent profiling for competitive advantage. In: *Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies of S&T Systems*, ed. by H.F. Moed, W. Glänzel, U. Schmoch (Springer, Dordrecht 2004) pp. 587–612
38.28 J.P. Courtial: A coword analysis of scientometrics, Scientometrics **31**(3), 251–260 (1994)
38.29 E. Bassecoulard, M. Zitt: Patents and publications. The lexical connection. In: *Handbook of Quantitative Science and Technology Research, The Use of Publication and Patent Statistics in Studies of S&T Systems*, ed. by H.F. Moed, W. Glänzel, U. Schmoch (Springer, Dordrecht 2004) pp. 665–694
38.30 L. Leydesdorff: The university-industry knowledge relationship: Analyzing patents and the science base of technologies, J. Am. Soc. Inform. Sci. Technol. **55**(11), 991–1001 (2004)
38.31 F. Janssens, J. Leta, W. Glänzel, B. De Moor: Towards mapping library and information science, Inform. Process. Manag. **42**(6), 1614–1642 (2006)
38.32 G. Salton: *Automatic Information Organization and Retrieval* (McGraw-Hill, New York 1968)
38.33 G. Salton, A. Wong, C.S. Yang: A vector space model for information retrieval, J. Am. Soc. Inform. Sci. **18**(11), 613–620 (1975)
38.34 G. Salton, M.J. McGill: *Introduction to Modern Information Retrieval* (McGraw-Hill, New York 1983)
38.35 M. Lennon, D.S. Pierce, B.D. Tarry, P. Willett: An evaluation of some conflation algorithms for information retrieval, J. Inform. Sci. **3**, 177–183 (1981)
38.36 D. Harman: How effective is suffixing?, J. Am. Soc. Inform. Sci. **42**, 7–15 (1991)
38.37 B. Krovetz: *Word Sense Disambiguation for Large Text Databases*, Ph.D. Thesis (Department of Computer Science, University of Massachusetts, Amherst 1995)
38.38 M.F. Porter: 'Snowball: A language for stemming algorithms, snowball.tartarus.org/texts/introduction.html (2001)
38.39 C.J. van Rijsbergen, S.E. Robertson, M.F. Porter: *New Models in Probabilistic Information Retrieval*, British Library Research and Development Report 5587 (British Library, London 1980)
38.40 M.F. Porter: An algorithm for suffix stripping, Program **14**(3), 130–137 (1980)
38.41 S. Dumais: Improving the retrieval of information from external sources, Behav. Res. Methods **23**(2), 229–236 (1991)
38.42 G. Salton, H. Wu: A term weighting model based on utility theory. In: *Information Retrieval Research*, ed. by R.N. Oddy, S.E. Robertson, C.J. van Rijsbergen, R.W. Williams (Butterworths, Boston 1981) pp. 9–22
38.43 C.E. Shannon: A note on the concept of entropy, Bell Syst. Tech. J. **27**(3), 379–423 (1948)
38.44 M.F. Moens: *Information Extraction: Algorithms and Prospects in a Retrieval Context*, The Information Retrieval Series, Vol. 21 (Springer, New York 2006)
38.45 T.K. Landauer, S.T. Dumais: A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge, Psychol. Rev. **104**(2), 211–240 (1997)
38.46 J.D. Carroll, P. Arabie: Multidimensional scaling, Annu. Rev. Psychol. **31**, 607–649 (1980)
38.47 K. Sparck Jones: *Automatic Keyword Classification for Information Retrieval* (Buttersworth, London 1971)
38.48 N. Jardin, C.J. van Rijsbergen: The use of hierarchic clustering in information retrieval, Inform. Storage Retr. **7**, 217–240 (1971)
38.49 P. Atherton, H. Borko: A test of factor-analytically derived automated classification method applied to descriptions of work and search requests of nuclear physicists, Report No. AIP/DRP 65-1, New York,

American Institute of Physics, Documentation Research Project; Report No. SP-1905, Santa Monica, California, System Development Corporation (1965), 15 p
38.50 H. Borko, M.D. Bemick: Automatic document classification, J. ACM **10**, 151–162 (1963)
38.51 P.G. Ossorio: Classification space: A multivariate procedure for automatic document indexing and retrieval, Multivar. Behav. Res. **1**, 419–524 (1966)
38.52 C. Eckart, G. Young: The approximation of one matrix by another of lower rank, Psychometrika **I**, 211–218 (1936)
38.53 M.W. Berry, Z. Drmac, E. Jessup: Matrices, vector spaces, and information retrieval, SIAM Rev. **41**, 335–362 (1999)
38.54 E. Jessup, J. Martin: Taking a new look at the latent semantic analysis approach to information retrieval. In: *Computational Information Retrieval*, ed. by M.W. Berry (SIAM, Philadelphia 2001) pp. 121–144
38.55 M. Lizza, F. Sartoretto: A comparative analysis of LSI strategies. In: *Computational Information Retrieval*, ed. by M.W. Berry (SIAM, Philadelphia 2001) pp. 121–144
38.56 M.W. Berry, M. Browne: *Understanding Search Engines: Mathematical Modeling and Text Retrieval* (Society for Industrial and Applied Mathematics, Philadelphia 1999)
38.57 D.M. Blei, A.Y. Ng, M.I. Jordan: Latent Dirichlet allocation, J. Mach. Learn. Res. **3**, 993–1022 (2003)
38.58 R. Baeza-Yates, B. Ribeiro-Neto: *Modern Information Retrieval* (Addison-Wesley, Wokingham 1999), Second edition published in 2011
38.59 D. Harman: An experimental study of the factors important in document ranking. In: *Assoc. Comput. Mach. 9th Conf. Res. Develop. Inform. Retr*, ed. by F. Rabbit (Association for Computing Machines, New York 1986)
38.60 P. Glenisson, W. Glänzel, F. Janssens, B. De Moor: Combining full text and bibliometric information in mapping scientific disciplines, Inform. Process. Manag. **41**, 1548–1572 (2005)
38.61 T. Magerman, B. Van Looy, B. Baesens, K. Debackere: *Assessment of Latent Semantic Analysis (LSA) Text Mining Algorithms for Large Scale Mapping of Patent and Scientific Publication Documents* (Faculty of Business and Economics, KU Leuven, Leuven 2011)
38.62 OECD: *A Framework for Biotechnology Statistics* (OECD, Paris 2005) pp. 29–32
38.63 X.H. Phan, L.M. Nguyen, S. Horiguchi: Learning to classify short and sparse text and web with hidden topics from large-scale data collections. In: *Proc. 17th Int. World Wide Web Conf. (WWW 2008), Beijing* (2008) pp. 91–100

Bart Van Looy

Faculty of Economics and Business
KU Leuven
Leuven, Belgium
bart.vanlooy@kuleuven.be

Bart Van Looy is Professor Innovation and Organization at the Department of Managerial Economics, Strategy and Innovation (MSI), Faculty of Economics and Business, KU Leuven. He is co-promotor of ECOOM (responsible for activities pertaining to technometrics/patent analysis) and a member of INCENTiM, a research division of KU Leuven Research and Development. His current research focuses on innovation processes (firm level) and regional innovation systems (entrepreneurial universities and science–technology interactions).

Tom Magerman

Faculty of Economics and Business
KU Leuven
Leuven, Belgium
tom.magerman@kuleuven.be

Tom Magerman is Lecturer at University College Leuven-Limburg (Applied Informatics) and Affiliated Researcher at the Faculty of Economics and Business, KU Leuven. He has a background in informatics and obtained a PhD in Applied Economics; he is active in the application of text and data mining techniques in the field of innovation studies (patents, publications, and science-technology interactions). Current research topics include record linkage, approximate string matching, natural language processing (NLP), and performance tuning for big data applications.

39. Application of Text-Analytics in Quantitative Study of Science and Technology

Samira Ranaei, Arho Suominen, Alan Porter, Tuomo Kässi

The quantitative study of science, technology and innovation (ST&I) has experienced significant growth with advancements in disciplines such as mathematics, computer science and information sciences. From the early studies utilizing the statistics method, graph theory, to citations or co-authorship, the state of the art in quantitative methods leverages natural language processing and machine learning. However, there is no unified methodological approach within the research community or a comprehensive understanding of how to exploit text-mining potentials to address ST&I research objectives. Therefore, this chapter intends to present the state of the art of text mining within the framework of ST&I. The major contribution of the chapter is twofold; first, it provides a review of the literature on how text mining extended the quantitative methods applied in ST&I and highlights major methodological challenges. Second, it discusses two hands-on detailed case studies on how to implement the text analytics routine.

39.1 Background

The quantitative study of science, technology and innovation (ST&I) is an independent subdiscipline of science and technology studies (STS) [39.1], and a research area that has been seen notable growth since the 1980s. This subfield of STS utilizes social science approaches to examine the relationships between ST&I via quantitative methods. Within the context of STS, ST&I refers specifically to the production and utilization of codified technical and scientific knowledge [39.2]. A wide range of qualitative approaches are available for ST&I research, for instance, laboratory ethnography [39.3] which utilizes systematic observation of participants in science and technology settings—in other words, observing or interviewing scientists and engineers. Other qualitative methods use interviews with expert panels, which is the approach found in the Delphi method [39.4]. Although qualitative research results are rich in detail and quite rigorous, they are not scalable. Therefore, if automatic tracing of technological development and innovation activities on a worldwide scale or country/organization level is desired, quantitative methods would be far more useful.

Quantitative ST&I approaches have been established as an independent stream of literature through the introduction of various methods and indicators enabling empirical research using patents, scientific papers [39.5–9] and other quantifiable sources. Analyses of empirical data with quantitative techniques increasingly contribute to ST&I endeavors. Mining research and development (R&D) electronic data sources [39.10–12] can provide firms with technical intelligence capabilities.

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_39

There is a long history of research related to the quantitative analysis of science and technology. The development of approaches for measuring scientific output dates back to as early as 1926, when *Lotka* [39.13] examined the scientific productivity of chemistry and physics researchers. Later, in 1949, Zipf's law, with its insight into word frequency versus rank frequency in natural language, provided a statistical basis for quantitative textual analysis. In 1965, the work of *Derek de Solla Price* [39.5] laid a strong foundation for the quantitative study of science and technology (S&T) using citation patterns. The creation of the Science Citation Index (SCI) by Eugene Garfield in 1965 was a turning point in quantitative STS, and since that time, numerous bibliometric methods have been developed based on statistics related to the production and distribution of documents rather than the textual content of the documents themselves; such methods include, for example, bibliographic coupling [39.14], citation [39.6] and co-citation analyses [39.7]. Later, co-word analysis [39.15] enabled examination of technical phrases extracted from the abstract, title or keywords. More recently, the proliferation of data sources and increasingly complex analysis in S&T management have led to the application of data-mining methodologies to textual data [39.16–22]. The trend in quantitative ST&I research of using increasingly complex methodologies is evidenced by the emergence of concepts such as *tech mining* [39.16], which applies text mining to the natural language of research documents with the aim of extracting valuable intelligence.

Despite the considerable attention given to quantitative methods in ST&I and the clear research advantages brought by text mining, there is thus far no consensus on a particular quantitative method, R&D or innovation indicator within the research community [39.1]. The quantitative study of ST&I has become distanced from the mainstream of qualitative ST&I and has become aggregated with fields like information science (informatics) [39.1]. Yet, there is a lack of comprehensive knowledge on how to fully exploit the potential of text mining to satisfy ST&I research objectives. Therefore, a major goal of this chapter is to present a state-of-the-art text-mining framework for application with ST&I research.

Narrowing our focus to the application of various text-mining techniques to ST&I data sources, the contribution of this chapter will be twofold: First, a systematic literature review covering text analytics techniques utilized in research articles in the ST&I research area is presented. The main objectives are to identify the types of text-mining methods that have been used to address ST&I research questions and to highlight the current challenges from a methodological perspective. Second, two hands-on case studies with detailed descriptions of the methodological process are presented to illustrate:

1. How to apply text mining to enhance a conventional patent search by capturing contextual information rather than using a Boolean search and identifying and discarding irrelevant documents
2. How to cluster scientific publications based on their content similarity and define the underlying topics.

For the purposes of hands-on application, the two case studies are complemented with Python programming codes to enable adoption of the methods by researchers and practitioners.

39.2 Literature Review on the Application of Text Mining

For literature review, the data collection process was started by formulating a search query on the Web of Science (WoS) database based on a list of terms including text mining, text analytics, document clustering or classification, machine learning and tech mining. The keywords were selected by the manual screening of recently published papers in terms of *authors' keywords* and *keyword plus* assigned by WoS. VantagePoint software was used in an iterative process for keyword consolidation. The first result list contained the WoS categories of computer science and linguistics as dominant clusters. Since this paper focuses on the application of text mining in ST&I, the search query was limited to articles classified under the WoS categories that gained the largest citations from ST&I methodological papers. A previously published literature review by *Martin* et al. [39.1] has identified highly cited research fields by papers in the quantitative study of science and technology [39.1, p. 1188]. The WoS categories selected for this study are similar to *Martin* et al. [39.1] as follows: *Business, management, economics, multidisciplinary science, social science interdisciplinary, operation research management science, and information science and library science, computer science multidisciplinary application and planning development*. To ensure the relevancy of the selected WoS category and articles, the analytics option of the WoS database was used to repeatedly screen top journals while selecting and excluding WoS categories. The titles and abstracts of the resulting 590 records were then manually screened to evaluate the relevancy of articles that employed text-mining techniques to address ST&I research questions.

As a result of the screening, a fine-grain sample of 154 papers from 18 journals was selected for the literature review process. The top three journals, namely, *Expert Systems with Applications*, *Technological Forecasting and Social Change* and *Scientometrics*, accounted for more than half of the 154 papers reviewed (Table 39.1). The annual distribution of the selected journal articles (Fig. 39.1) shows a steady growth of papers using text mining-based methods under the ST&I research theme. The year 2006 was selected as a starting point, because the predecessor to this Springer Handbook, titled [39.23]: *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems* was published in 2005.

To examine field-specific topics, keyword co-occurrences based on Pearson correlation between the author-assigned keywords were analyzed in VantagePoint software. Figure 39.2 shows two distinct clusters. Keywords on the left represent text-mining and machine learning methods and concepts, while the cluster on the right describes ST&I subfields and scientometric tools. The relationship between the clusters shed light on ST&I research areas that take advantage of text-mining techniques. For instance, patent mapping has a strong correlation with the subject-action-object text-based method (Fig. 39.2), which indicates that many authors carried out patent mapping using a text analytics approach. To examine correlations in more detail, the 154 papers were reviewed based on the concepts presented in Fig. 39.3. The papers were examined and annotated in terms of three criteria: their major practical or theoretical contributions to ST&I, the data sources used to support the research objectives, and the principles of the text mining-based methodology utilized:

- *Contribution to ST&I*: This criterion divided the papers in terms of their contribution to consideration of ST&I research objectives or development of research tools in operationalization of the research targets. As can be seen in Fig. 39.3, many research tools have benefited from the application of text-mining methods. The arrow suggests these developed research tools can be used to fulfill research objectives in multiple subfields of ST&I; for instance, patent mapping in operationalization of technology forecasting activities, technology transfer and detection of technological opportunities. Science mapping and quantitative measurement of multidisciplinarity have been utilized in research evaluation projects. The literature-based discovery method and measurement of S&T interaction have been employed in research papers utilizing technological opportunities analysis (TOA).
- *Data sources*: The increasing availability of electronic documents, the enormous amount of information stored in ST&I data sources and rapid growth of social web documents have opened up new oppor-

Table 39.1 Top six journals

Journal title	Number of papers
Expert Systems with Applications	42
Technological Forecasting and Social Change	23
Scientometrics	22
Information Processing & Management	17
Journal of the American Society for Information Science and Technology	13
Journal of Information Science	6

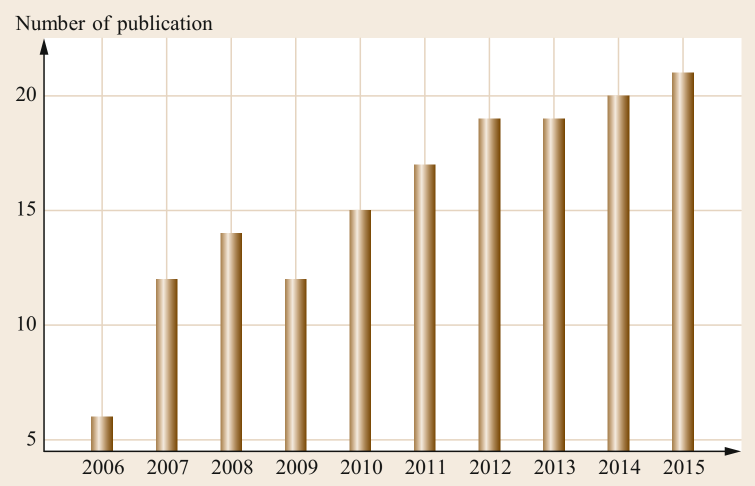


Fig. 39.1 Annual number of published scientific articles utilizing text-mining techniques in the field of ST&I (2006–2015)

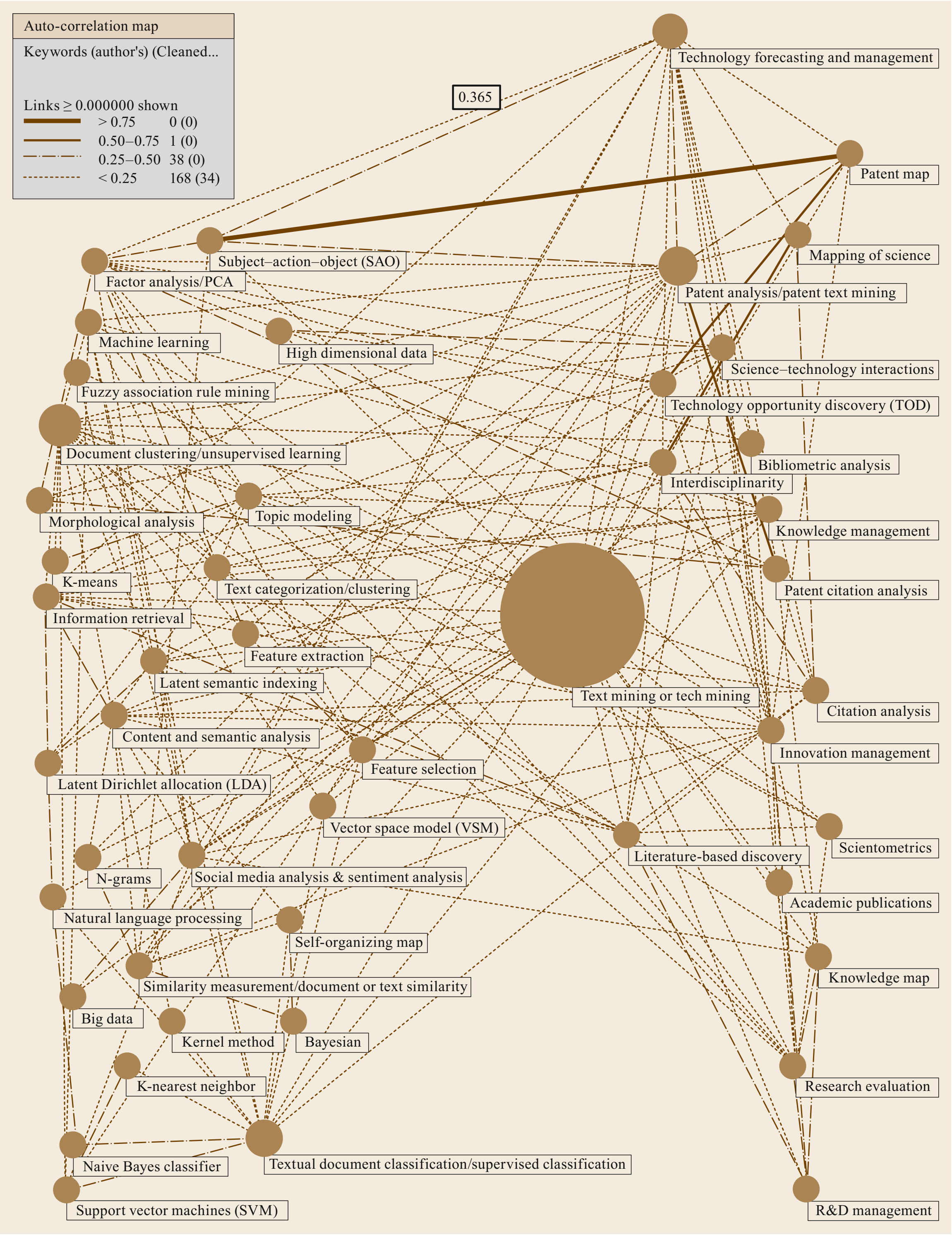


Fig. 39.2 Auto-correlation map of author-assigned keywords

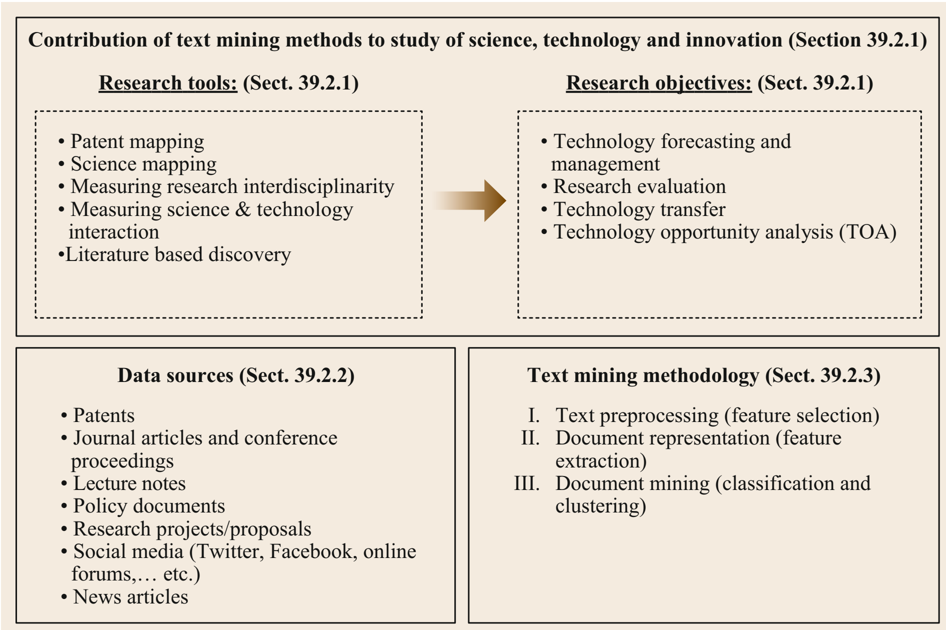


Fig. 39.3 Conceptual components of the literature review

tunities for social scientists to analyze and use these data in their research projects. Sect. 39.2.2 discusses the data sources used by the reviewed papers.

- *Methodological approach*: While the main role of ST&I data sources is data storage and responding to users' search queries, a substantial stream of research is directed largely toward the development of methodologies to extract data, and to process and reveal useful information. Text mining has been applied in ST&I research as a complementary tool to bibliometric approaches. *Leopold* et al. [39.24] presents an overview of text mining and its relation with natural language processing and machine learning, and how the techniques work together to tackle research problems in science and technology studies. As a continuation to [39.24], Sect. 39.2.3 evaluates ST&I papers from the perspective of text-based methodologies employed during the last decade.

39.2.1 Contribution of Text Mining Methods to Study of Science, Technology and Innovation

Research Objectives

The final labeling of 154 papers resulted in 112 documents directly contributing to the enhancement of research tools or fulfilling the ST&I research objectives. The contribution of the remaining 44 records can be summarized as purely methodological; proposing a new text analytics method or the improvement to existing approaches, which is tested on ST&I-related data sources (e. g., news articles, social media, scientific publications).

A number of research subfields were identified in the selected papers. The largest subfield is related to technological forecasting (TF) and management (Fig. 39.3), and approximately one-third of the collected papers examined subjects in this area. In the past, forecasting of technological development relied on qualitative methodologies (e. g., expert opinion and the Delphi method), and text mining-based methods provided data-driven evidence to back up forecasting practices in ST&I. According to *Porter* et al. [39.25], technology forecasting is a broad concept that intersects with several research activities: technology monitoring [39.26], technology assessment [39.27–29], technology road mapping [39.30–33], foresight research [39.34, 35] and impact analysis [39.36].

Technology monitoring refers to the process of identification and evaluation of technological development trends that are critical to a firm's competitive position. Departing from the established citation-based analysis, *Lee* et al. [39.26] built a mathematical framework based on patent keywords that can overcome the limitations of conventional patent citation maps. The major drawback of citation-based methods for most technology forecasting purposes is that only citing-cited patents are considered and it usually takes some time for a patent document to receive a citation so that it would be included in the analysis. Thus, a clear limitation to using patent citation-based methods is that patents without citations are omitted.

In the context of technology assessment, *Britt* et al. [39.28] suggested that the overhead costs associated with technological readiness could be reduced by utilizing text-mining techniques to exploit the technology readiness reports [39.28] and to semantically cluster textual documents. Automatic interpretation of technology readiness documents can allow firms to keep abreast of rapid changes in the market. *Kostoff* et al. [39.27] classified scientific publications based on content similarity as part of a methodology used for country-level technology assessment. Similarly, *Alencar* et al. [39.29] assessed the nanotechnology development efforts of Japan, the USA and Europe based on textual content analysis of patent abstracts.

Technology road mapping (TRM) is a strategic product planning tool first introduced to the academic community in 1987 by the Motorola company [39.37]. The approach links strategic management goals to product development at the firm level. Conventional TRM methods emphasize qualitative and expert knowledge rather than utilizing quantitative information. However, a recent biometrics analysis and review of TRM research shows growing use of quantitative approaches in operationalization of TRM [39.38]. The application of text mining enables the exploitation of a large quantity of important information sources, from product manuals to patent documents, that might otherwise be left untapped [39.31, 33]. The extracted key information consists mainly of the most important keywords that frequently occur in the documents which serve as input for the examination of the relationships among the documents. *Yoon* and *Phaal* [39.32] identified a series of data-mining and text-mining techniques that can be specifically used in TRM to yield important information from raw textual data. *Lee* et al. [39.39] formulated a technology-driven road map by combining text-mining techniques with citation network analysis to extract information from patent data.

Foresight research is another core avenue of research focused on supporting decision-making processes related to ST&I policies and activities at the country level. Foresight is operationalized by mainly qualitative approaches, such as communication and discussion with stakeholders, systematic interviews and expert panels. The combination of conventional foresight research methods with quantitative methods has shown promising research output [39.34]. A case study was carried out by the Center for Management and Strategic Studies (CGEE) in Brazil [39.34] that made use of text-mining techniques in combination with expert opinion to analyze nanotechnology R&D activities. The data-driven output of their foresight practice assisted funding agencies in the decision-making process.

Research evaluation is the process of measuring the impact of academic publications in the scientific community, and research evaluation commonly uses bibliometric approaches. An increasing number of researchers have questioned the validity of raw citation counts as a proxy for research evaluation and journal and author ranking [39.40], which has led to consideration of alternative text-based approaches [39.41–46]. *Liu* et al. [39.41] enhanced the performance of scientific publication ranking by combining full text citation analysis with document content analysis. The analysis of the papers' textual information characterizes the citation relationships between the papers. As a result, the published papers could be ranked based on their scientific contribution to the associated topic, rather than the number of received citations. In comparison with traditional citation-based journal ranking approaches, the new journal ranking system favors new publications with significant contributions to the field from not well-known authors who have not yet received many citations.

The detection of core documents within a particular circle of the scholars is an important topic in research evaluation. The concept of core documents was introduced by *Henry Small* in 1973 using a classical co-citation method [39.7]; later, the detection of core documents was carried out based on bibliographic coupling by Glänzel and Czerwon in 1996. Recently, *Glänzel* and *Thijs* [39.42] extended the notion of core documents by merging bibliographic coupling with a lexicon similarity component. In the case of missing citation links, the new hybrid method [39.42] is able to connect documents with textual similarity and calculate term frequencies to label the citation-based clusters. Investigating research trends, *Sunikka* and *Bragge* [39.43] applied a research profiling approach [39.39] built upon text mining to distinguish two new niche research streams in marketing and management research.

When tracing research trends, the detection of leading scientists' names (eponyms) is as important as identifying their research output. *Cabanac* [39.44] performed a semi-automated text mining of 821 *Scientometrics* papers that led to the unveiling of numerous already known and emerging eponyms. Other work utilizing text mining for research evaluation purposes includes [39.45], where the impact of European Union (EU)-funded research in social sciences and humanities on EU policies was evaluated by analyzing the content of research papers and policy documents, and [39.46], where, based on the assumption that research quality is dependent on the educational system, institutions' research capability was assessed by analyzing the contextual relationship between lecture material and research output.

Technology transfer refers to the dissemination of technologies from the original producers to a group of firms or non-commercial actors (e. g., universities) that are seeking to exploit external knowledge and innovation. The underlying tasks in successful technology transfer include the detection and acquisition of high-value technologies and/or commercialization of a firm's technological output. In [39.47], a text-mining approach based on TRIZ (theory of inventive problem solving) theory that enables the automatic detection of a promising solution from patent texts was employed to address a technological problem related to floating wind turbines. The function-based patent analysis method proposed in [39.48] provided firms with knowledge about the potential applications of a particular technology in different industries. It was argued in [39.48] that there is a lack of consistency between industries in the use of terminology to describe different technologies with similar functionality. First, the terminologies used to express a functionality in patents were extracted. Second, the authors in [39.48] linked these terminologies to similar functions in different industries. This linkage process may lead to the detection of additional applications for a technology. Other work using text analytics to ascertain technological and business opportunities from patent documents include [39.49–52].

Research Tools

A number of scientometrics studies have designed research tools based on text analytics methods that fulfill multiple research objectives in ST&I (Fig. 39.3). As a subfield of TF activity, such tools attempt to detect emerging technologies (ET), make new scientific discoveries or identify hot subjects in social media and news. Classical tools for the detection of ETs are bibliometric analysis and citation-based approaches; this section focuses on methods utilizing text mining or combining text mining with other approaches.

The available established ST&I-related data sources are designed primarily for data storage and information retrieval and not for the purpose of scientometrics research. For instance, major ST&I data sources like patent and publication databases are designed to facilitate data storage and easy information retrieval through a classification systems and to be organized by patent examiners and librarians, respectively. It is thus a challenging task to link database classification schemes to specific products [39.53] or industries [39.54] to retrieve relevant information for the detection of ETs. Extraction of topics from the abundance of user-generated information available in social media or news streams requires advanced text-mining methods to be used. Therefore, numerous studies have explored alternative quantitative methods to circumvent the limitations of traditional approaches for data retrieval from ST&I data sources.

The science, technology and innovation policy (STIP) research group at the Georgia Institute of Technology devised a bibliometrics-based search strategy [39.55] using text-mining software (VantagePoint) to capture the emergence of nanotechnology research output. In addition to the formulation of a search query for the bibliometric analysis of nanotechnology, *Kostoff* et al. [39.56] identified institutions and countries with mutual interest in nanotechnology research by using similar terminologies extracted from the literature as a proxy to map them into the corresponding institution or country. Newly emerging science and technology (NEST) has been studied by a combination of bibliometric analysis and text mining in support of technology management and policy [39.57]. Literature-based discovery (LBD), whose foremost advocate is *Ronald Kostoff*, refers to a particular type of text mining that seeks to identify nontrivial information from a large body of documents, and the approach has been applied to a number of real-world case studies, for example, to discover resolutions for water purification and the treatment of human diseases [39.58–61].

Several studies [39.62, 63] argue that current text-mining methods for the detection of emerging research topics overlook the novelty of subjects by focusing only on frequency measures. *Tu* and *Seng* [39.62] proposed two indices based on publication time, journal volume and subject frequency to capture the novelty of the research topic, and *Tang* et al. [39.63] proposed a blended index by integrating a text similarity measure with word frequency at the sentence level for real-time novelty mining. To uncover latent research topics in the e-commerce field, *Cheng* et al. [39.64] utilized a chance discovery method grounded in artificial intelligence that evaluates data from the perspectives of term frequency and association links. Other work exploring research trends includes research by *Delen* and *Crossland* [39.65], who described a semi-automated text-mining method to detect research trends, and *Yang* et al. [39.66] proposed a link-bridged topic model (LBT) that combines linkage between documents defined by co-citation and a document's textual content.

Patent mapping is a patent analysis tool that has been used for many different purposes in technology management, and various text-mining methods have greatly facilitated its implementation. An extensive procedure for automating the whole patent mapping process by incorporating text-mining techniques for summary extraction, text segmentation, term association, document cluster generation and topic identification is described in [39.67, 68]. The manual implementation of these tasks requires an analyst with knowledge and expertise in the disciplines of information retrieval, natural language processing, technology domain knowledge, and business intelligence. According to [39.67], the creation and maintenance of a patent map on *carbon nanotube* technology from 100 patent documents takes five analysts more than a month. The proposed automated patent mapping tool [39.67] facilitates the detection of previously unknown and useful patterns from large patent text repositories. *Trappey* [39.69] studied patent summary extraction by combining ontology and the concept clustering approach to capture general knowledge and the core meaning of patents in a given domain. Other studies utilizing text-mining methods for the construction of patent maps, for instance *Lee* et al. and *Son* et al. [39.39, 70], have proposed a keyword-based patent map for discovering undeveloped or unexplored technological fields and significant rare keywords [39.71] from patent databases. A text mining-based patent map has been developed to support the merger and acquisition (M&A) decision process by evaluating firms from the technological perspective [39.72]. Keyword-based patent maps show the topology of the target technology, from which researchers can learn about core technological information. In patent maps, the relationship between core technologies within a technology domain is lacking. A recent patent map proposed by *Choi* [39.73] combined a keyword-based patent map with a community network approach that informs researchers about the interaction between essential technology elements, highlighted as important keywords, in corresponding technology fields.

Retrieval of relevant patent documents is an essential step prior to any mapping and patent analysis task. The dominant approach for finding similar patents or retrieving relevant patents for a technology domain is construction of complex search queries based on international patent classification (IPC) classes and/or keywords. However, such patent searches are challenging as regards linking IPC classes to an industry [39.54] or product-level analysis [39.74]. The IPC classification scheme has been designed to ease the process of storing documents by examiners and does not directly facilitate the retrieval process. Furthermore, patent classification schemes are somewhat subjective, as they are based on examiners' judgment [39.75], and IPC classes are often considered too broad or detailed to be applied directly to a specific area of technology or research interest. Recent studies have introduced semantic patent retrieval systems as a complement to keyword-based [39.76] methods or in combination with bibliometric coupling [39.77]. *Venugopalan* and *Rai* [39.78] applied a machine-learning algorithm to classify patents into relevant/irrelevant groups based on extracted linguistics features.

Science mapping represents a family of tools that have been widely applied by the scientometrics community for research evaluation, measuring research interdisciplinarity and visualizing knowledge structure. Several studies have attempted to use text analytics to execute subject-based journal classification schemes [39.79, 80], build keyword-based knowledge maps that illustrate key information from a research proposal repository [39.81], and compensate the absence of semantic linkages in existing word co-occurrence-based knowledge maps [39.82]. Mapping of scientific disciplines facilitates the identification of boundaries and measurement of interdisciplinarity. Interdisciplinary research (IDR) is a mode of research that integrates concepts, theories and techniques from different disciplines [39.83]. IDR has a crucial role in pushing scientific boundaries forward, as it seeks to solve research problems whose solutions lie beyond a single body of knowledge [39.84]. Topic mapping and text analytics of research proposals have provided useful new information about IDR [39.85].

The debate regarding the interaction between science and technology (S&T) has a long history in scientometrics, dating back to 1963, when Toynbee compared S&T to dance partners [39.86]. The S&T relationship dynamic is central to policy-oriented research and innovation studies [39.87]. S&T interaction has also been measured by a set of diverse bibliometric indicators pertaining to scientific and technological advancement (patents and publications) [39.88, 89]. The feasibility of applying text-based semantic analysis to illustrate the topical overlap between patents and publications has recently been examined [39.90]. A study by *Magerman* et al. [39.90] shows that using the latent semantic indexing technique is a valuable approach in detection of content similarity between patent–paper pairs, but may not yield an accurate outcome in small data sets. Another study led by *Magerman* [39.91] incorporates semantic text analysis with patents' forward citations to detect patent–paper pairs in the field of biotechnology. The results showed that scholars who are involved in both scientific publishing and technological patenting have a larger scientific impact than their peers whose activities are restricted to scientific publishing. A Taiwanese research group's exploration of biofuel patent papers using a text-based clustering approach led to the identification of potential scientific and technological applications for micro-algal biofuels [39.92].

39.2.2 Data Type

Over two-thirds of the papers in our sample utilized either patents or scientific publications as major data sources (approximately 107 records). The remaining records exploited other sources of knowledge to provide evidence for ST&I research. For instance, content analysis by the National Science Foundation (NSF) provided new information for measuring interdisciplinarity [39.85], novel technological solutions, and new ideas derived from descriptions of research projects awarded grants from the National Institute of Standards and Technology (NIST) in the United States [39.93] or the Ministry of Defense in Germany [39.36]. For capturing interest in an emerging research topic from outside the academic domain, data sources other than patents and scientific publications have been explored using text analytics.

Text sentiment analysis [39.94], a new frontier in text mining, involves the analysis of social media data. A topic discovery system presented by [39.95] aimed to reveal implicit knowledge present in news streams. *Ma* et al. [39.96] used machine-learning techniques to predict the future popularity of Twitter topics based on keywords with hashtags. *Lu* et al. [39.97] utilized text analytics to detect hot topics in online health communities that might provide knowledge about patients' needs and interests. The exploration of market changes via online news articles by applying text analytics [39.51] enables companies to identify business intelligence factors.

39.2.3 Text-Mining Approaches Used in the Reviewed Literature

The core functionality of the text-mining approach lies in the identification of concept co-occurrence patterns across document collections [39.98]. In practice, text mining utilizes algorithmic approaches to identify distributions, frequency sets and associations of concepts at an inter-document level to illustrate the structure and relationships of the concepts as reflected in the corpus [39.98]. The main goal of text mining is to derive implicit knowledge from textual information by applying an array of methods from statistics, natural language processing and machine learning. Text-mining algorithms require a mathematical representation of text documents; thus, a wide range of text extraction and transformation approaches are available for use. Almost all the reviewed articles followed all or at least part of the three major text preprocessing and modeling phases shown in Fig. 39.4.

Step A

The main objective in the first phase is extraction of valuable informative terms (features) from the text. The reviewed articles [39.42, 50, 62, 67, 99] applied various feature extraction approaches, which can be categorized as morphological analysis, syntactical analysis or semantic analysis. Morphological analysis was utilized as a text preprocessing routine by most of the papers. The first step in the process is tokenization, which reduces sentences to words (tokens) and removes punctuation. Next, stop-words such as `a`, `the` or `and` that carry no semantic meaning are eliminated. This is followed by stemming, a linguistic normalization technique in which a token is reduced to its root (stem) by removal of derivational suffixes. For instance, all variations of the verb `applying`, `applied`, `applies` are transformed to `appl`. In some cases, the final word's appearance might not be recognized by the user or the text analytics algorithm. As a remedy, one or more sophisticated morphological methods known as lemmatization can be used to convert the words of a sentence to their dictionary base.

Continuing with the above example, lemmatization would return `appl` to a common form of `applying`, `applied`, `applies`. N-gram is another common method that allows analysis to be conducted at single word (unigram) or phrase level, i. e., phrases that consists of two or three words (bigram, trigram, etc.). Syntactical analysis provides knowledge of the grammatical structure of the sentences. The meaning of a sentence is easier to interpret once the correct grammatical and semantic information has been defined. For instance, in part-of-speech (POS) tagging, a sentence

Text mining steps

A: Feature extraction	B: Feature selection	C: Document mining
• Morphological analysis • Syntactical analysis • Semantic analysis	• Vector space model (VSM) or Bag of worlds (BOW) • Sequence-based representation • Dimension reduction methods	• Clustering • Classification

Fig. 39.4 Methodological steps in text mining

is often annotated by its noun, verb, adjective, adverb, and so on. POS is used in [39.100] to make sense of a bilingual English-Chinese document collection as part of a proposed business intelligence system.

Semantic analysis incorporates lexicons and dictionaries to categorize words with similar meanings. Semantic lexicons are the backbone of sentiment analysis studies, which attempt to measure the emotion of information posted on social media in news articles [39.101], online forum discussions [39.102], customer reviews, tweets and Facebook statuses or comments [39.103]. *Li* and *Wu* [39.102] integrated text mining with sentiment analysis to evaluate the emotional polarity of an online sport forum using the HowNet (http://www.keenage.com) lexicon. Other studies have applied sentiment analysis to social media to capture information about social aspects of a topic, for example, for research objectives such as competitor analysis within an industry [39.103] and opinion mining [39.104].

Step B

Feature selection is the second phase in text processing, where the documents are represented based on a fixed informative subset of terms by the removal of redundant information. The vector space model (VSM) [39.105], which is grounded on the singular-value decomposition (SVD) method, is a document representation approach found in many of the articles reviewed [39.30, 39, 67, 77, 91, 106, 107]. The VSM represents documents as weighted high-dimensional vectors, where the dimensions pertain to individual features such as words or phrases. When only words are used, the model is called a bag of words (BoW). The weight assigned to vectors is often calculated by standard weighting schemes such as simple term frequency or term frequency inverse document frequency (TF-IDF). TF-IDF measures are widely used to show the high-frequency words across the collection or rare terms. For instance, a study by [39.71] was able to extract particularly rare terms for a patent document retrieval process by combining TF-IDF with a term association metric [39.45, 67–69, 91, 108]. Despite being a common word-ranking system in information retrieval, the TF-IDF function does not consider the length of documents. The text available in social media news, patent abstracts and full-text articles differs in terms of size. To achieve more robust representation, recent studies [39.109, 110] have applied the Best Match 25 (BM25) weighting scheme, which includes a document length normalization component.

The high dimensionality of VSM models poses challenges to examination of the subject of the document. Additionally, many features are weighted as a zero value, which means they do not appear in the term vectors. The existence of many vectors with zero values is known as a sparsity problem. The dominant dimensionality reduction approach is principal component analysis (PCA) (or singular-value decomposition), which transforms high-dimensional data to a low-rank estimation of sparse matrices. PCA captures vectors with larger variances, since the high-variance components contain more information. PCA was applied by *Zhang* et al. [39.99] for ST&I text analysis to consolidate topical content from patent/publication text. Other reviewed papers that have applied PCA either for dimension reduction purposes or for grouping of similar words include [39.39, 57, 68, 90, 97, 111–113].

Another common approach in feature selection is representing documents based on a term's syntactical order. The POS tagging process can be used to produce a sequence-based representation of the documents. Recently, the subject-action-object (SAO) structure, which is composed of subject (noun), action (verb), and object (noun), has attracted research attention [39.47, 48, 114]. The SAO method reflects the key concepts and structural relationships between a sentence's components. Using the SAO approach, valuable findings in patent abstracts could be uncovered by extracting the provided solution to the technological problem [39.47]. For example, *Choi* et al. [39.114] illustrated the relationship between product components, various technologies and their particular functions from patents related to the proton-exchange membrane (PEM) fuel cell. Other scholars have applied SAO to identify the potential applications of a particular technology [39.48].

Step C

In the last step, the text analytics and machine-learning approaches will be employed for semantically organizing documents extracted from ST&I databases. The challenge in data analysis at the document level falls into two main tasks: document classification and clustering. In the machine learning literature, the former task refers to the categorization of documents in a supervised manner based on a set of predefined patterns, which is called the training data set. The latter task is unsupervised classification or clustering of documents on the basis of their similarity without a priori knowledge from training data. Within our 154 sample articles, the proportion of papers that applied clustering algorithms is larger than those employing classification approaches. It may be assumed that clustering is more appealing due to the effort, time and cost associated with training a supervised algorithm.

The different supervised algorithms utilized by the reviewed articles are shown in Fig. 39.5 (for detailed mathematical background on the methods, please refer to [39.24]). Papers that used commercialized software

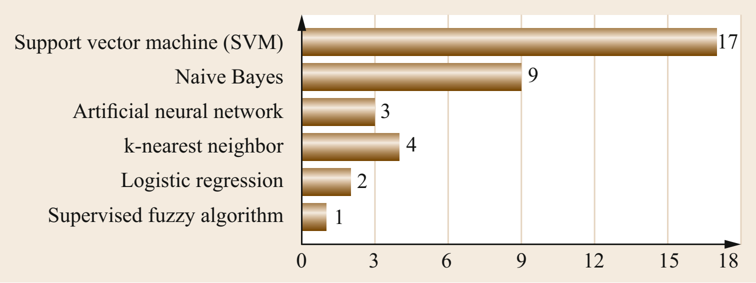


Fig. 39.5 Supervised classification algorithms utilized by reviewed papers

or feature selection methods or did not reveal the algorithm used have been excluded.

As can be seen in Fig. 39.5, the most frequently applied classifier is a relatively new learning approach, the support vector machine (SVM), introduced by Vapnik in 1995 to solve two-class pattern recognition problems. SVM also performs well on high-dimensional data, and has been employed, for instance, to detect patents relevant to solar photovoltaics technology [39.78] and to establish a cross-language patent retrieval system for Japanese-English patents [39.115]. The case study discussed in Sect. 39.3.1 employs the SVM classification algorithm to execute an automatic patent classification task.

The second commonly used method, probabilistic naive Bayes (NB), classifies documents based on an assumption that the terms (document features) are all equally important and are independent of each other. An NB algorithm has been applied to categorize tweets based on their sentiments (positive or negative discussion) [39.116] and for organization of news articles in the Portuguese language [39.35]. A bundle of methods like SVM, NB, *k*-nearest neighborhood (a distance-based algorithm) and logistic regression are applied in [39.96] at different stages of their methodology to predict and classify future popular topics on Twitter, which might be valuable information from marketing and economic perspectives.

The most common document clustering approach (Fig. 39.6) is classic PCA [39.99], which also represents the other reviewed articles that used its substance methods, such as factor analysis [39.56] and multidimensional scaling (MDS) [39.33]. In addition to document clustering, MDS can be used for dimension reduction purposes. However, PCA methods suffer from excessive information loss when pruning the data dimensions; moreover, they cannot account for correlated words within the given lexicon of the corpus. In other words, count-based methods that rely on merely the co-occurrence of words are not very accurate in document clustering tasks due to not being able to account for polysemy (words with multiple meaning) and synonymy (multiple words with similar meaning). As a possible remedy, latent semantic indexing (LSI) [39.79, 117], which includes the context (document) of the words, has been proposed. For instance, the two keywords `cell` and `electrode` are related to each other as a part of fuel cell technology, but may not co-occur many times. LSI as a context-based clustering method matches documents with context similarity, based on the probability of neighboring words in a particular document, rather than matching the keywords. The disadvantage of the LSI function is the absence of a solid probabilistic foundation.

The probabilistic latent semantic indexing (PLSI) method proposed by *Hofmann* [39.118] was a significant step forward in document clustering methods, as it provided a probabilistic structure at the word level as an alternative to LSI [39.66]. The PLSI model draws each word of a document from a mixture model

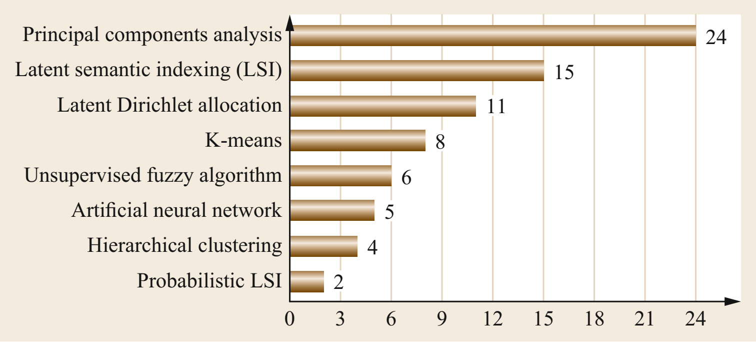


Fig. 39.6 Unsupervised clustering algorithms utilized by reviewed papers

specified via a multidimensional random variable. The mixture model represents the *topic*. Therefore, each word originates from a single topic, and different words of one document can be drawn from various topics. Even though PLSI is a valuable contribution to the text clustering field, it lacks a probabilistic model at the document level, because documents in PLSI are represented as a list of numbers, with no generative probabilistic model for the numbers. The limitations of PLSI at the document level cause problems such as over-fitting, as the number of parameters grows linearly with the size of the corpus.

The latent Dirichlet allocation (LDA) method [39.119] overcomes the limitations of PLSI and provides probabilistic models for both documents and words [39.85, 120]. LDA is a predictive model that draws latent topics from textual data. In LDA, documents are represented as a random mixture of latent topics and each topic is based on the distribution of the words. Another family of predictive models is the artificial neural network [39.121]. The remaining papers used the following clustering methods: *k*-means from partitioning methods [39.32], hierarchal clustering [39.81] and unsupervised fuzzy algorithms [39.27].

In general, all the above text mining-based techniques for document representation, clustering and classification are dependent in some way or another on distributional semantic models (DSMs). The term DSM dates back to the 1991 study by *Miller* and *Charles* [39.122], in which semantically similar words are placed near each other since they tend to have a similar contextual distribution. For example, vector space models (VSM) are count-based models [39.123] where semantically similar terms are embedded in a close neighbor in a vector.

Count-based methods compute the statistics of how often a word co-occurs with its neighbor words in a large text corpus, and then map these count-statistics down to a small, dense vector for each word. Recent studies [39.124, 125] show that predictive models can outperform count-based models in the natural language processing context, as they directly predict a word from its neighbor terms. Despite the shift toward predictive models, very few studies have applied them in the ST&I field to solve problems with science mapping [39.126] or clustering patent documents [39.78]. Motivated by this gap, the case study in Sect. 39.3.2 below considers the science mapping task using a predictive LDA model.

39.3 Case Studies

This section offers step-by-step guidance to utilizing text mining and machine learning, using Python packages, for automatic patent classification and clustering of scientific publications based on their content. The automatic patent classification example refers to a situation where a set of patent documents is available, for example, patents related to fuel cell electric vehicles (FCEV), and researchers wish to collect similar patents. The clustering of scientific publications example illustrates how research themes can be derived from a set of scientific publications on a broad subject. In this case, too, fuel cell technology, which is a multipurpose technology, is used as an example.

39.3.1 Case Study 1: Automatic Patent Document Classification

Using IPC codes to retrieve relevant patent documents has been shown to have limited value when studying electric vehicle (EV) technology [39.74]. *Pilkington* et al. [39.74] investigated the technological development of EVs using the patent class *B60L11*, which returned a significant number of irrelevant patents not necessarily related to electric automobiles. *Pilkington* [39.74] emphasizes that a clear boundary is required between generic patents related to electric device technologies and automotive-oriented patents. This may be achieved by using an archive of reliable keywords [39.127, 128]. However, a keyword-based Boolean search carries the risk of missing documents that lack the listed keywords. Another challenge is inconsistency among the terminologies used by patent applicants, inventors, researchers and attorneys. Phrases and terms become outdated at some point as new concepts, innovative materials and processes emerge. In addition, such database searches are based on the match of exact wording, which is merely a means of finding phrases without contextual meaning. Additionally, a lack of familiarity with the area of technology for which the patent data are being gathered can complicate construction of an exhaustive keyword list. For instance, a patent search query using `electric` and `vehicle` in an attempt to search for `electric vehicle`-related patented technology would produce patents related to any type of vehicle that uses an electric current [39.129]. The high number of irrelevant patents reduces the reliability of the search outcome.

For the purpose of this experiment, this case study was designed with the aim of using text mining to recognize patents relevant to the vehicle industry by capturing contextual information within the documents. The goal was to differentiate patents related to automotive engineering with a specific focus on fuel cell electric vehicles (FCEV) from irrelevant documents within a complex data set. The test set of 218 744 patent documents was gathered from the PATSTAT (EPO Worldwide Patent Statistical Database) database using the search terms `car`, `vehicle` and `automobile` for US-granted patent abstracts between the years 2005 and 2014. The patent search created a heterogeneous test set of technologies from automotive to medical applications.

The case study used the support vector machine (SVM) learning model to distinguish FCEV patents from non-FCEV patents. SVM is a powerful two-class pattern recognition classifier; in other words, it is able to answer yes or no questions, which makes it a suitable method for categorizing patents as relevant or irrelevant. Prior to the classification task, the SVM needs to be trained by a predefined labeled data set, called a training set. The training set formed in this study included 1374 records split into relevant (positive) and irrelevant (negative) groups. The positive group was collected based on a random selection of 700 patents related to fuel cell electric vehicles (FCEV) gathered using a cooperative patent classification (CPC) class (Y02T90/34). The CPC classification scheme is an extended version of IPC jointly developed by the European Patent Office (EPO) and the United States Patent and Trademark Office (USPTO). The main objective of CPC is to provide a harmonized classification system and improve patent searches through the provision of more detailed technology classes. It should be noted that at the time of this research, the Y02T class related to *Climate Change Mitigation Technologies* related to *Transportation* was not complete, which means that not all relevant FCEV patents could be retrieved relying merely on the CPC code. The FCEV patent abstract descriptions included information related to the electric motor, internal combustion engine (ICE), hybrid, plug-in and battery electric vehicle technologies. The negative group in the training set included inventions irrelevant to the vehicle industry, for example, patents in the areas of textiles, musical instruments and speech recognition.

Text Preprocessing

The abstract section of the training and test sets needs to be preprocessed before executing the SVM classification task. The Python code uses Pandas packages to import the patent abstract stored in a comma-separated values (CSV) file (Fig. 39.7). The built-in text processing uses functions from Scikit-learn (an open-source Python library that implements a range of machine learning, text processing and visualization tasks) and the Natural Language Toolkit (NLTK), which is the leading Python platform for human language analytics. Python libraries were used as the main platforms to conduct the text-mining steps. The full implementation of preprocessing and SVM classification was done in iPython. For the purpose of demonstration, the following snapshots illustrate the steps in detail. The first step is importing all required Python packages and the patent abstracts (Fig. 39.7).

The vector space model (VSM) was adopted for document representation using the feature extraction function from the Scikit-learn package (Fig. 39.8). The sentences from the document's abstract were transformed to lowercase and broken down into words. The reduction of sentences to keywords or phrases is called tokenization and is operationalized by the NLTK Python package. Phrases with combinations of two or three words (bigrams or trigrams) with the minimum frequency of *one* are included in the analysis.

The next preprocessing step is pruning words with very high values. Words that occur in more than half of the data collection have limited differentiating value in classification tasks. The cutoff value for removal of words was set as words occurring in more than 50% of the documents. Once the documents have been transformed into a compact representation of term vectors, the terms need to be weighted across the document collection to indicate their importance. The TFidfVectorizer Python function was used to assign weights to the term vectors.

```
#import csv, NLTK and scikit-learn packages
import sklearn, nltk
import as pandas as pd

#import patent abstracts from csv file using pandas package
train_data = pd.read_csv ("filename.csv")
train_labels= pd.read_csv ("filename.csv")
```

Fig. 39.7 Setting up the required Python packages and importing patent abstracts

```
#formulate the tokenization function and store tokens as a list
def tokenize(text):
tokens = nltk.word_tokenize(text)

#import TFidfVectorizer from scikitit-learn package
from sklearn.feature_extraction.text import TfidfVectorizer
#operationalise the vector space model and text-preprocessing
vectorizer = TfidfVectorizer(lowercase = True,
analyzer = 'word', preprocessor = None, tokenizer = tokenize,
ngram_range = (1, 2),  #consider bigrams and trigrams
min_df = 1,          #pruning word frequency threshold
max_df = 0.5,
stop_words = 'english', #enforce English stopword
use_idf = True)

#apply the VSM model to train-set and test-set
train_vectors = vectorizer.fit_transform(train_data)
test_vectors = vectorizer.transform(test_data)
```

Fig. 39.8 Implementation of SVM classification using the Scikit-learn Python program

SVM Implementation

Implementation of SVM classification comprises two steps (Fig. 39.9); the SVM classifier first needs to be trained based on the labels *relevant* and *irrelevant*, and then the trained algorithm is applied to predict relevant and irrelevant documents in the test set.

Results Evaluation

The classification task aimed to assign patent documents in the test set to the FCEV-relevant category. To evaluate the SVM performance, the test set was labeled based on the relevancy of the associated IPC code to FCEV technologies. The utilized IPC codes were retrieved based on the three major technology components of FCEV: *fuel cell technology* (H01M486, H01M496, H01M824, H01M1208), *brush-less* or *electric motor* (H02K19, H02K49) and *electric/battery-supplied vehicle* (B60L710, B60L722, B60L8, B60L9, H02J7). The labeling process of the test set resulted in 3447 documents out of 218 077 related to FCEV technology, and the remaining documents were annotated as other. This means that by using the list of relevant IPCs, only 3447 patents from the initial collection were recognized as patents relevant to FCEV.

Taking IPC classes as a baseline, true positive (TP) and true negative (TN) are relevant documents that the SVM predicted correctly and falsely, respectively. False positive (FP) and false negative (FN) are those patents defined by IPC as irrelevant to FCEV technology, where the latter is predicted to be relevant to FCEV technology. The SVM classification performed poorly (about 33%) calculated by the evaluation measures of precision (TP/TP + FP), recall (TP/TP + FN) and *F*-score (2TP/2TP + FP + FN) (Table 39.2).

The FN category comprises those patents that the SVM classifier predicted as being relevant to FCEV technology, but they were labeled as *other*. Qualitative screening of the FN category (4970 patent document titles and abstracts) indicated that about 3231 documents were relevant to FCEV technology components. During the qualitative screening, the labels of 3231 records in the FN category were updated and categorized under the *relevant* class.

The second round of evaluations is based on new FN values, and results are presented in Table 39.3. The implication is that IPC was not an ideal baseline for assessment of the SVM performance. The qualitative review of the documents showed that the SVM is

Table 39.2 SVM classification performance considering IPC as a baseline

Evaluation	Precision	Recall	*F*-score
Evaluation based on IPC as a baseline	0.483	0.251	0.33

```
#train SVM classifier with the vectorized train-set data
classifier_rbf = svm.SVC()
classifier_rbf.fit(train_vectors, train_labels)

#apply the trained SVM on the test-set collection to predit
prediction_rbf = classifier_rbf.predict(test_vectors)
print(classification_report(test_labels, prediction_rbf))
```

Fig. 39.9 Implementation of SVM classification algorithm

Table 39.3 SVM classification performance against qualitative labeling of the test set

Evaluation	Precision	Recall	F-score
Evaluation based on qualitative screening	0.733	0.737	0.735

capable of correctly predicting the right class for a substantial portion of the documents, with an F-measure of 0.735. It can be argued that the algorithm performance has been improved using qualitative labeling of the test set, while it still shows about 27% error.

This case study provided an automatic approach for a patent retrieval process. Utilization of the machine learning classifier can be considered as a complementary technique for a patent search task. The classifier can be useful for situations where researchers possess a set of patents and need to collect more patent documents with similar topics. It should be also noted that the precision of algorithm classification performance highly depends on the quality of the training set.

39.3.2 Case Study 2: Science Mapping

In the second case study, fuel cell publications are examined and research themes are identified using an unsupervised classifier. Fuel cells are seen as having great potential for use in non-combustion engine vehicles. Fuel cells were invented in 1838 [39.130], and a number of trials to develop a device for practical applications have been carried out. Expensive materials and the low conversion efficiency have discouraged the development of fuel cell technology and, for decades, fuel cells remained an uninteresting technological option. Only in the past 25 years have fuel cells taken leaps forward in maturity. The family of fuel cells offers a variety of solutions, ranging from large stationary applications to small milliwatt systems [39.131]. The most significant expectations had been expressed for portable, stationary and transportation solutions.

For this case study, data were downloaded from published English journals and conference articles from the ISI Web of Science using a search query for `fuel cell` or `fuel cells` in the title, abstract, descriptors or identifiers of a publication. The search query is not limited to any time window. This search resulted in a data set of 75 479 articles, from which the bibliographical data were downloaded. Figure 39.10 gives the count of publications from 1996 until 2016 and shows the growth of the publications has been significant. A clear increase in publications is apparent from the late 1990s. The data were searched in mid-2016; thus, data for 2016 are incomplete and cannot be used as a yearly total value.

This case study uses an LDA algorithm to cluster the publications. LDA, unlike supervised or reinforced models, creates an outcome relying solely on its for-

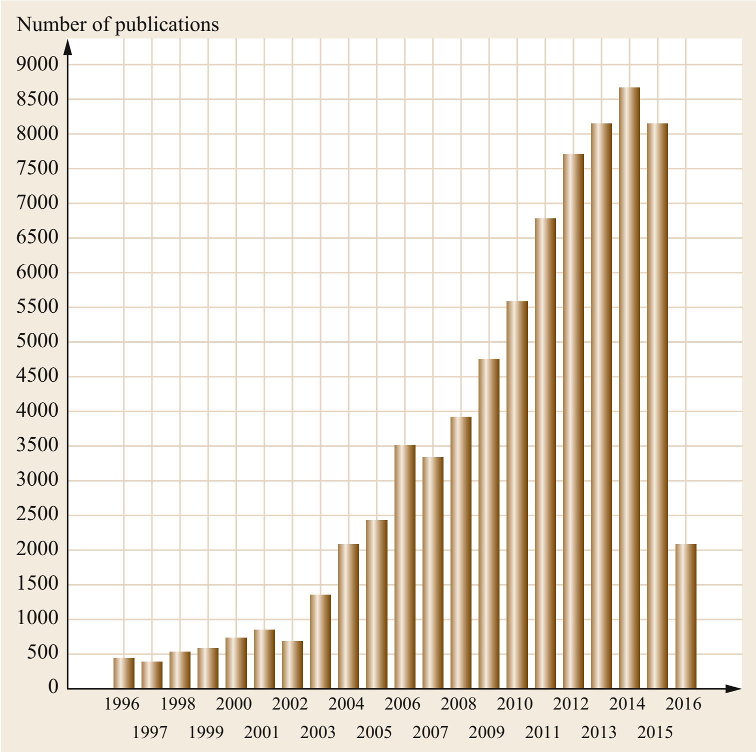


Fig. 39.10 Number of scientific publications related to fuel cell technology

mal framework. Several studies have investigated the applicability of LDA to uncover patterns from text collections drawn from different sources, for example, scientific publications [39.126, 132] and patents [39.111, 133, 134]. LDA is a soft classification algorithm, which means that the algorithm classifies inputs by giving a probability of the item belonging to each cluster. As researchers often strive for hard classifications, there is a need to reduce the dimensionality of the LDA results to one class. Approaches for this task can be simple, such as using the most probable class as the class an item is assigned to, or more elaborate, for example, dimension reduction using a modularity algorithm [39.135] or PCA. Simple hard classification selection of the most probable topic loses a significant portion of the LDA results, whereas the other approaches integrate the soft classification as a part of their results.

Analyzing scientometrics data with topic models, specifically LDA, has attracted significant research interest in recent years. *Yau* et al. [39.132] analyzed the effectiveness of different approaches for topic modeling, including LDA, and considered specifically the algorithms' precision and recall on a human-labeled document set. *Suominen* and *Toivanen* [39.126] used LDA as a method to create a science map. In the patent domain, *Venugopalan* and *Rai* [39.78] applied topic modeling to patent data looking at knowledge spillovers. In the photovoltaic context, the authors were able to reduce the dimensionality of the input data, highlighting an aggregation of the patents' knowledge content. *Lee* et al. [39.134] analyzed technological convergence using LDA to identify keywords from the document corpus. *Suominen* et al. [39.133] used patent data to structure the knowledge portfolio of telecommunication companies. Using full-text patent data, the study structured the technological portfolios of the selected companies and created a forecast of the technology pathways of the sample companies.

Text Preprocessing

In the preprocessing phase (Fig. 39.11), the Python package Pandas was used to import the flat file and to identify whether fields had varying data types.

```
#Import, read data and create variables
import pandas as pd
from nltk.stem import WordNetLemmatizer
wnl = WordNetLemmatizer()
df=pd.read_csv(filename, sep="\t", header=0, encoding="latin-1")

#Creating a list of abstracts, identifiers and publication years
abList = df['AB'].tolist(), utList = df['UT'].tolist()
pyList = df['PY'].tolist()
#Pre-processing-Creating a list to store pre-processed abstracts
abPrecessed=[] , idProcessed =[]
counter=0

#For loop to pre-process abstracts
for line in abList:
        #Here we exclude absracts shorter than 100 characters
        #Remove item from utList and pyList if abstract removed
        if len(line)<100:
                continue
        else:
                #Removing stopwords given in a user created
                data= ' '.join([word for word in line.split()
                        if word not in stopwords])
                #terms containing a number removed
                data = ' '.join(s for s in data.split()
                        if not any(c.isdigit() for c in s))
                #lemmatization
                data = " ".join([wnl.lemmatize(i)
                        for i in data.split(" ")])

                abProcessed.append(data)
                idProcessed.append(utList[counter])
                pyProcessed.append(pyList[counter])
        counter += 1
```

Fig. 39.11 Preprocessing for LDA analysis

The WordNet lemmatizer was imported from the NLTK package for word consolidation. For stop-words, the code relies on a user-provided list. Although a stop-word implementation could be provided through NLTK, there is some case specificity regarding what should be removed from the input data. In addition, as the objective is to use the LDA with science abstracts, research has shown that excessive word removal prior to analysis can diminish the quality of the results [39.132].

As the source data is abstract, aggressive preprocessing can in practice yield extremely short input per document (e. g., a 100-word abstract with excessive preprocessing can be cut to one with little or nothing to analyze). To avoid including abstracts with little or no content, the code includes an `if` statement excluding abstracts smaller than 100 characters. This limit is arbitrary and can be changed dependent on the input used. The main goal here is to exclude documents with no or extremely short abstracts.

After preprocessing, the Gensim package was utilized to run the LDA algorithm (Fig. 39.12). First, the code imports the required packages; thereafter, the code uses the list created in the preprocessing stage to create a corpus tokenized for analysis. At this stage, Gensim offers an option to implement stop-word removal, but as this was implemented at the beginning of the analysis, it was not done here. Gensim also includes an option to remove words with either high or low occurrence rates in the corpus. In the code, the `limitBelow` variable removes the words that appear fewer than `limitBelow` times. The variable `limitExtremes` keeps only the first `limitExtremes` most frequent words. The `limitAbove` variable is a fractional filter to keep words which are contained in no more than `limitAbove` documents. These variables are given as parameters to `filter_extremes`.

After creating the corpus and before running the analysis, the LDA requires several further inputs in addition to the data for the model. One of the most significant inputs is selection of the number of topics created by the model. There is currently no consensus on the most practical method for assigning the number of topics. Some researchers have claimed that a trial-and-error method of testing a different number of topics

```
from gensim import corpora, models, similarities, matutils
from gensim.models import ldamodel
from itertools import izip as zip
import numpy as np
import scipy.stats as stats
#Variables
limitBelow = 20, limitAbove = 0.9, limitExtremes = 50000,
valueMin =1,valueMax =100, valueStep =1

#create a dictionary
dictionary = corpora.Dictionary(line.lower().split()
                    for line in abProcessed)
# remove stop words and words that appear only once
once_ids = [tokenid for tokenid, docfreq in
                        dictionary.dfs.items() if docfreq <= 1]
dictionary.filter_tokens(once_ids)
# remove gaps in id sequence after words that were removed
dictionary.compactify()
#Filter extremes
dictionary.filter_extremes(no_below=limitBelow, keep_n = limitExtremes)

class MyCorpus(object):
def __iter__(self):
 for line in abProcessed:
  yield dictionary.doc2bow(line.lower().split())
FCcorpus = MyCorpus()
# Parameters are the corpus and dictionary for the data,
# lower and upper bound of the topic evaluation and
# max and step between topics to be evaluated
lVector = np.array([sum(cnt for _, cnt in doc)
                        for doc in FCcorpus])
kl = klDivergence(FCcorpus,dictionary, lVector, valueMin,
                        valueMax, valueStep)
```

Fig. 39.12 Creating a corpus and evaluating KL divergence

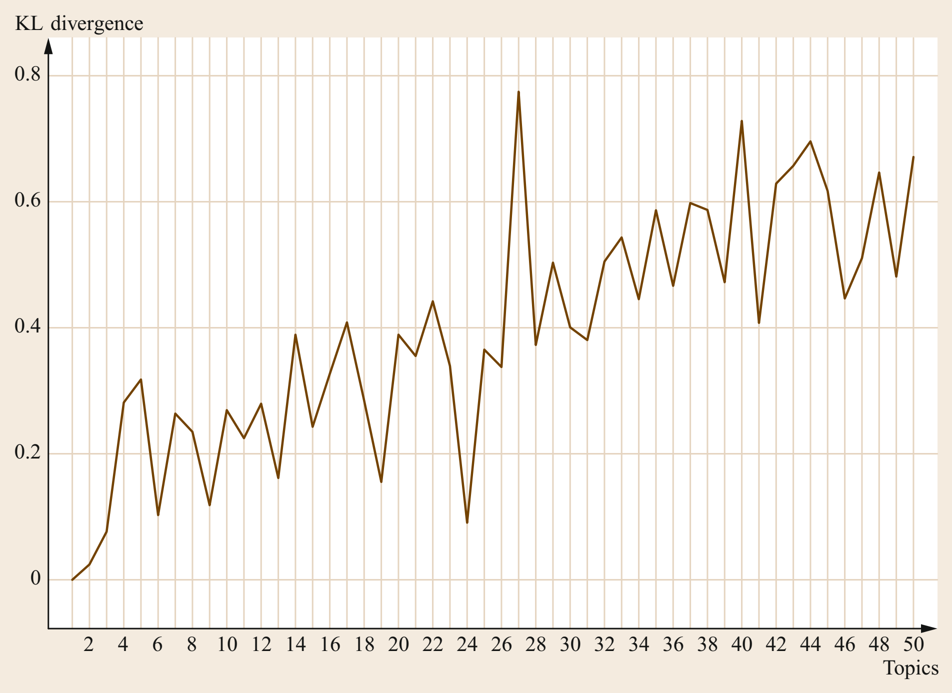


Fig. 39.13 Calculation of KL divergence to identify optimum number of topics

```
lda = models.ldamodel.LdaModel(FCcorpus, id2word=dictionary,
num_topics=24)
lda_corpus = lda[corpus]
lda.save(LDA_model)
```

Fig. 39.14 Running of the final topic value for the LDA model

with given input data will produce results that are most convenient for human interpretation [39.136]. However, a number of other mathematical approaches have also been proposed, such as using Kullback–Leibler (KL) divergence to estimate the input [39.137]. In this study, the Python code proposed by [39.137] was used. KL divergence was implemented to estimate the number of topics in the constructed fuel cell corpus.

Figure 39.13 shows the plot of values returned by the function KL divergence. As seen in Fig. 39.13, estimation of the number of topics requires human intervention and simply taking the smallest value of the series is not sufficient. Even if the researcher has a relatively narrow window of expected topics, automating evaluation of a KL divergence plot can be challenging. In the case at hand, the number of topics selected for the analysis was 24 topics. This selection was based on the sharp decrease and value of the KL divergence at 24 topics.

Topic Modeling Results with the LDA Algorithm

Based on the assessment of the number of topics, the final model was created using the LDA model (Fig. 39.14). Gensim offers a convenient means of running a model with three parameters: a corpus, a dictionary and the number of topics. The model can be saved for later use, for example, to infer new documents against the same model.

The model completes with a two-result data set: the probability distribution of documents being classified to a topic and the probability distribution of words being associated with a topic. At this point, the researcher has several avenues through which to approach the results. The documents-topic probabilities can be further used to compare the proportion of topics or study the linkages between topics. The word-topic probabilities can be used as a source for an auto-labeling algorithm or plotting a word-cloud enabling qualitative interpretation of topic content.

To describe the topics, word clouds were implemented using the R word-clouds package, which enables visualization of the topics in a format convenient for human interpretation. Figure 39.15 shows two examples of word clouds from the 22 topics created. The word clouds show the thematic differences in the topics created. For example, Fig. 39.15a describes a vehicle-

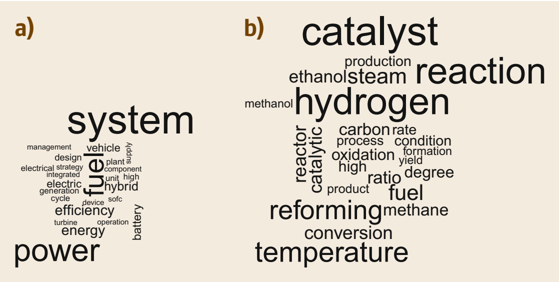


Fig. 39.15a,b Selected topics from the 24 LDA topics: **(a)** topic 1, **(b)** topic 7

related thematic topic; it is also focused on hydrogen as a fuel, storing fuel in a vehicle and the efficiency of the vehicle system.

The document topic probabilities were retrieved using the call `lda[doc]` (Fig. 39.16). This call retrieves top topic probabilities, omitting extremely small probabilities. In practice, this creates a directed vertex from a document to a topic that has a weight that is the probability of a document belonging to a topic. For example, the code in Fig. 39.16 shows the probability distribution of first document in the corpus related to four major topics (topic numbers 7, 10, 12 and 16).

All probabilities creating a bipartite network where nodes are topics and documents and edges are links between documents and topics were extracted, which allowed reduction of the dimensionality of the data in two ways. First, the data were reduced such that the documents were hard-classified to only one class using the modularity algorithm [39.135]. This creates new classes, called communities, that contain both topics and documents. Second, the dimensionality of the data was reduced by transforming the bipartite graph to a one-mode projection, allowing the links between topics to be understood. This transformation can be done for the bipartite graph of topics and for the communities' documents. If done with the latter, it should be noted that a community can include multiple topics.

The bipartite network created from the data consisted of 69 942 nodes and 342 029 edges (Fig. 39.17). Of the nodes, 24 were topics and the rest were documents that remained in the analysis after preprocessing had removed the short documents. Running the modularity algorithm at a resolution of 1.0 produced a hard classification with nine communities. Table 39.4 shows the percentage distribution of nodes to each community, as well as the topics that were hard-classified to a specific community. The average community size is 11.11% ($s = 3.12\%$, $N = 9$), which suggests a relatively equal distribution of nodes in each community. A qualitative assessment of the linked topics, based on the word-clouds, resulted in the labeling given in Table 39.4. Although subjective, the labels give an indication of the content in each community and enable further discovery of the topic content.

The dimensionality of the bipartite network can be reduced to show interactions between the topics. This requires transformation of the bipartite network to a one-mode representation that shows the topic nodes and calculates their interaction based on the shared probabilities between the topics. A bipartite network illustrates the LDA model as a topic-to-topic representation. Figure 39.17 shows the interactions between the topic nodes and reduces the complexity of fuel cell technology development to a few clusters, namely temperature issues in orange, PEM fuel cells in lilac, nanostructures and materials in blue, and hydrogen fuel systems in green.

```
for doc in FCcorpus:
 print(lda[doc])
 [(7, 0.035872179268285884),(10, 0.26206672752133908),
 (12, 0.50151704492601012),(16, 0.050475303440253319)]
 ...
```

Fig. 39.16 Print document topic probabilities

Table 39.4 Communities, the topics embedded to a community and the share of nodes in each community

Community	Topics	Node (%)	Label
1	Topics 1, 12 and 20	12.51	Fuel systems
2	Topics 5 and 8	13.08	Hydrogen fuel systems
3	Topics 9, 10 and 13	8.25	Fuel cell stack
4	Topics 3, 4, 11 and 21	11.14	Nanostructures
5	Topics 6, 14 and 15	12.18	Small fuel cells
6	Topics 7 and 16	9.32	Catalyst and electrode
7	Topic 18	6.32	Solid oxide fuel cell
8	Topics 2 and 23	9.47	Temperature issues
9	Topics 17, 19, 22 and 24	17.72	PEM fuel cells

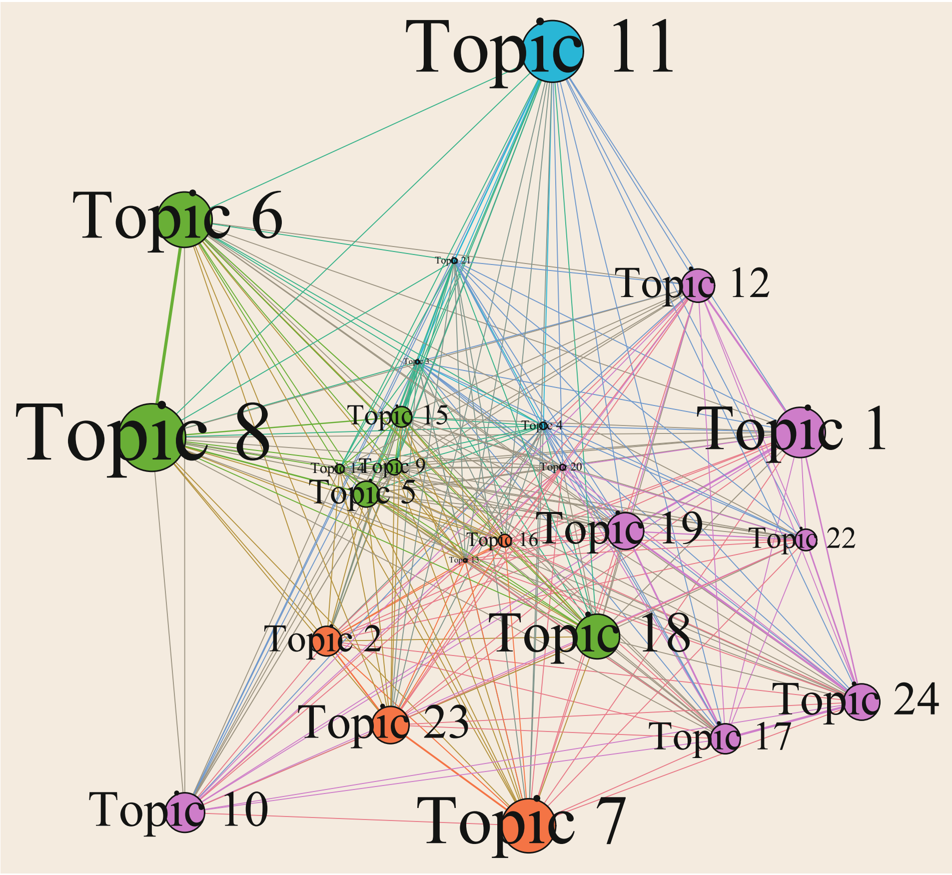


Fig. 39.17 One-mode representation of the bipartite graph created from the LDA. *Node size* is based on hard-clustered documents in each topic. *Color* is based on modularity run for the one-mode representation

39.4 Discussion and Conclusion

The major contribution of this chapter is twofold: First, it has provided a summary of how text mining extends the range of quantitative methods applicable to ST&I and highlighted major methodological challenges. A set of 154 articles was examined based on their major theoretical or practical contribution to ST&I, data source utilized, and text-mining technique applied. Second, it has illustrated hands-on detailed case studies showing how a text analytics routine can be implemented.

The literature review shows that considerable progress has been made in terms of incorporating text analytics methods in the development of scientometrics tools to address ST&I research objectives. Scientometrics research tools using citation-based methods that have been partially coupled with or fully developed based on text analytics include science mapping, patent mapping, interdisciplinary research indicators, measurement of S&T interaction and literature-based discovery. Among the ST&I research subfields, the operationalization of technology forecasting and management, research evaluation, technology transfer and detection of technological opportunities has benefited widely from the utilization of text mining. In addition to patents and scientific papers, which are known to be major sources of codified technological and scientific knowledge, other novel data sources (e. g., social media, news streams, policy documents, research proposals) have been exploited over the last decade.

The literature review summarized the text-mining routines practiced by scholars, starting from alternative techniques of text extraction and preprocessing to document modeling. Implementation of the first two text-mining steps described as feature extraction and feature selection (Fig. 39.4) are necessary to process and present the text in a mathematical format suitable for machine-learning algorithms. In general, all document modeling problems can be classified as either supervised document classification or unsupervised document clustering. The first case study presented in this chapter illustrated a document classification challenge where similar patent documents must be collected and classified to a particular known sample of patents, in this particular case related to FCEV technology. The second case addressed the document clustering issue, that is, identifying the underlying topics of an unknown document set. Science mapping is an example of document clustering, as it involves the grouping of scientific publications with similar research themes. Open-source Python libraries were used to implement the case studies, and detailed guidance of their use is provided.

Broadly speaking, analysis of high-dimensional natural language is a complex task due to three prominent challenges: text ambiguity, redundancy and the absence of semantic relationships between words or documents. Ambiguity appears when multiple words have similar meaning (polysemy), redundancy refers to situations where several words share similar meaning (synonymy), and the lack of semantic links occurs when the context of the extracted words is neglected. Dominant text analytics methods such as vector space models (VSM) or principal component analysis (PCA) (i. e., count-based methods) may not be able to overcome all three challenges for document modeling tasks. However, PCA and VSM perform well for word-level analysis whose aim is to extract important keywords in the corpus. The advancement of count-based approaches toward the family of methods with superior performance known as context-based predictive methods [39.123] is well reflected in the literature. While count-based models remain the dominant approach, an increasing number of papers in recent years have applied predictive models to execute science mapping, technology forecasting activities and research evaluation. Acknowledging the advantages of predictive models would affect how future studies in ST&I can shape their research design.

Overall, the analysis supports the contention of an evolution of quantitative ST&I analysis methods from elementary count-based methods and linkage metrics to complex natural language analysis [39.138]. Current natural language analysis studies have laid a firm foundation for future work, and the studies enable a robust understanding of methodological options and limitations when analyzing complex science and technology data. The two presented case studies illustrate that researchers already have a fairly robust armamentarium of natural language and machine-learning-based analysis methods to answer complex ST&I-related research questions.

References

39.1 B.R. Martin, P. Nightingale, A. Yegros-Yegros: Science and technology studies: Exploring the knowledge base, Res. Policy **41**(7), 1182–1204 (2012)

39.2 M.B. Jensen, B. Johnson, E. Lorenz, B.Å. Lundvall: Forms of knowledge and modes of innovation, Res. Policy **36**(5), 680–693 (2007)

39.3 S. Traweek, A. Kernan: Beamtimes and lifetimes: The world of high energy physicists, Phys. Today **43**, 68 (1990)

39.4 H.A. Linstone, T. Murray: *The Delphi Method: Techniques and Applications*, Vol. 29 (Addison Wesley, Reading 1975)

39.5 D.J. De Solla Price: *Little Science, Big Science and Beyond* (Columbia Univ. Press, New York 1965)

39.6 E. Garfield, I.H. Sher, R.J. Torpie: *The Use of Citation Data in Writing the History of Science* (Institute for Scientific Information, Philadelphia 1964)

39.7 H. Small: Co-citation in the scientific literature: A new measure of the relationship between two documents, J. Assoc. Inf. Sci. Technol. **24**(4), 265–269 (1973)

39.8 F. Narin, E. Noma, R. Perry: Patents as indicators of corporate technological strength, Res. Policy **16**(2–4), 143–155 (1987)

39.9 Z. Griliches: Patent statistics as economic indicators: A survey, J. Econ. Lit. **28**, 1661–1707 (1990)

39.10 A.L. Porter, N.C. Newman: Mining external R&D, Technovation **31**(4), 171–176 (2011)

39.11 T.U. Daim, G. Rueda, H. Martin, P. Gerdsri: Forecasting emerging technologies: Use of bibliometrics and patent analysis, Technol. Forecast. Soc. Change **73**(8), 981–1012 (2006)

39.12 P. Érdi, K. Makovi, Z. Somogyvári, K. Strandburg, J. Tobochnik, P. Volf, L. Zalányi: Prediction of emerging technologies based on analysis of the US patent citation network, Scientometrics **95**(1), 225–242 (2013)

39.13 A.J. Lotka: The frequency distribution of scientific productivity, J. Wash. Acad. Sci. **16**(12), 317–323 (1926)

39.14 M.M. Kessler: Bibliographic coupling between scientific papers, J. Assoc. Inf. Sci. Technol. **14**(1), 10–25 (1963)

39.15 M. Callon, J.-P. Courtial, W.A. Turner, S. Bauin: From translations to problematic networks: An introduction to co-word analysis, Soc. Sci. Inf. **22**(2), 191–235 (1983)

39.16 A.L. Porter, S.W. Cunningham: *Tech Mining: Exploiting New Technologies for Competitive Advantage* (Wiley, San Francisco 2005)

39.17 P. Losiewicz, D.W. Oard, R.N. Kostoff: Textual data mining to support science and technology management, J. Intell. Inf. Syst. **15**(2), 99–119 (2000)

39.18 B. Yoon, Y. Park: A text-mining-based patent network: Analytical tool for high-technology trend, J. High Technol. Manag. Res. **15**(1), 37–50 (2004)
39.19 A. Hotho, A. Nürnberger, G. Paaß: A brief survey of text mining, LDV Forum **20**(1), 19–62 (2005)
39.20 F. Madani, C. Weber: The evolution of patent mining: Applying bibliometrics analysis and keyword network analysis, World Pat. Inf. **46**, 32–48 (2016)
39.21 A. Abbas, L. Zhang, S.U. Khan: A literature review on the state-of-the-art in patent analysis, World Pat. Inf. **37**, 3–13 (2014)
39.22 L. Zhang, L. Li, T. Li: Patent mining: A survey, SIGKDD Explor. **16**(2), 1–19 (2015)
39.23 H.F. Moed, W. Glänzel, U. Schmoch: *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems* (Springer, Dordrecht 2005)
39.24 E. Leopold, M. May, G. Paass: Data mining and text mining for science and technology research. In: *Handbook of Quantitative Science and Technology Research*, ed. by H.F. Moed, W. Glänzel, U. Schmoch (Springer, Dordrecht 2005) pp. 187–213
39.25 A.L. Porter, S.W. Cunningham, J. Banks, R. Thomas, M. Thomas, R. Frederick: *Forecasting and Management of Technology*, 2nd edn. (Wiley, New York 2011)
39.26 C. Lee, J. Jeon, Y. Park: Monitoring trends of technological changes based on the dynamic patent lattice: A modified formal concept analysis approach, Technol. Forecast. Soc. Change **78**(4), 690–702 (2011)
39.27 R.N. Kostoff, A. DelRio, H.D. Cortes, C. Smith, A. Smith, C. Wagner, L. Leydesdorff, G. Karypis, G. Malpohl, R. Tshiteya: Clustering methodologies for identifying country core competencies, J. Inf. Sci. **33**(331), 21–40 (2007)
39.28 B.L. Britt, M.W. Berry, M. Browne, M.A. Merrell, J. Kolpack: Document classification techniques for automated technology readiness level analysis, J. Am. Soc. Inf. Sci. Technol. **59**(4), 675–680 (2008)
39.29 M.S.M. Alencar, A.L. Porter, A.M.S. Antunes: Nanopatenting patterns in relation to product life cycle, Technol. Forecast. Soc. Change **74**(9), 1661–1680 (2007)
39.30 B. Yoon, R. Phaal, D. Probert: Morphology analysis for technology roadmapping: Application of text mining, R&D Manag. **38**(1), 51–68 (2007)
39.31 S. Choi, H. Kim, J. Yoon, K. Kim, J.Y. Lee: An SAO-based text-mining approach for technology roadmapping using patent information, R&D Manag. **43**(1), 52–74 (2013)
39.32 B. Yoon, R. Phaal: Structuring technological information for technology roadmapping: Data mining approach, Technol. Anal. Strateg. Manag. **25**(9), 1119–1137 (2013)
39.33 S. Lee, S. Lee, H. Seol, Y. Park: Using patent information for designing new product and technology: Keyword based technology roadmapping, R&D Manag. **38**(2), 169–188 (2008)
39.34 M. de Miranda Santo, G.M. Coelho, D.M. dos Santos, L.F. Filho: Text mining as a valuable tool in foresight exercises: A study on nanotechnology, Technol. Forecast. Soc. Change **73**(8), 1013–1027 (2006)
39.35 P.S.M. Nizer, J.C. Nievola: Predicting published news effect in the Brazilian stock market, Expert Syst. Appl. **39**(12), 10674–10680 (2012)
39.36 D. Thorleuchter, D. Van den Poel: Quantitative cross impact analysis with latent semantic indexing, Expert Syst. Appl. **41**(2), 406–411 (2014)
39.37 C.H. Willyard, C.W. McClees: Motorola's technology roadmap process, Res. Manag. **30**(5), 13–19 (1987)
39.38 N. Gerdsri, A. Kongthon, R.S. Vatananan: Mapping the knowledge evolution and professional network in the field of technology roadmapping: A bibliometric analysis, Technol. Anal. Strateg. Manag. **25**(4), 403–422 (2013)
39.39 S. Lee, B. Yoon, Y. Park: An approach to discovering new technology opportunities: Keyword-based patent map approach, Technovation **29**(6/7), 481–497 (2009)
39.40 M.H. MacRoberts, B.R. MacRoberts: Problems of citation analysis: A study of uncited and seldom-cited influences, J. Am. Soc. Inf. Sci. Technol. **61**(1), 1–12 (2010)
39.41 X. Liu, J. Zhang, C. Guo: Full-text citation analysis: A new method to enhance scholarly networks, J. Am. Soc. Inf. Sci. Technol. **64**(9), 1852–1863 (2013)
39.42 W. Glänzel, B. Thijs: Using 'core documents' for the representation of clusters and topics, Scientometrics **88**(1), 297–309 (2011)
39.43 A. Sunikka, J. Bragge: Applying text-mining to personalization and customization research literature? Who, what and where?, Expert Syst. Appl. **39**(11), 10049–10058 (2012)
39.44 G. Cabanac: Extracting and quantifying eponyms in full-text articles, Scientometrics **98**(3), 1631–1645 (2014)
39.45 A. Patrick, D. Mazière, M.M. Van Hulle: A clustering study of a 7000 EU document inventory using MDS and SOM, Expert Syst. Appl. **38**(7), 8835–8849 (2011)
39.46 H. Lee, J. Kwak, M. Song, C.O. Kim: Coherence analysis of research and education using topic modeling, Scientometrics **102**, 1119–1137 (2014)
39.47 H. Park, J.J. Ree, K. Kim: Identification of promising patents for technology transfers using TRIZ evolution trends, Expert Syst. Appl. **40**(2), 736–743 (2013)
39.48 H. Park, J. Yoon, K. Kim: Using function-based patent analysis to identify potential application areas of technology for technology transfer, Expert Syst. Appl. **40**(13), 5260–5265 (2013)
39.49 H. Seol, S. Lee, C. Kim: Identifying new business areas using patent information: A DEA and text mining approach, Expert Syst. Appl. **38**(4), 2933–2941 (2011)
39.50 B. Yoon, I. Park, B.-Y. Coh: Exploring technological opportunities by linking technology and products: Application of morphology analysis and text mining, Technol. Forecast. Soc. Change **86**, 287–303 (2014)

39.51 W. Chung: BizPro: Extracting and categorizing business intelligence factors from textual news articles, Int. J. Inf. Manag. **34**(2), 272–284 (2014)

39.52 J. Yoon, H. Park, W. Seo, J.-M. Lee, B.-Y. Coh, J. Kim: Technology opportunity discovery (TOD) from existing technologies and products: A function-based TOD framework, Technol. Forecast. Soc. Change **100**, 153–167 (2015)

39.53 A. Pilkington, R. Dyerson: Innovation in disruptive regulatory environments: A patent study of electric vehicle technology development, Eur. J. Innov. Manag. **9**(1), 79–91 (2006)

39.54 U. Schmoch: *Concept of a Technology Classification for Country Comparison*, Final Report to the World Intellectual Property Organization (Fraunhofer ISI, Karlsruhe 2008)

39.55 S.K. Arora, A.L. Porter, J. Youtie, P. Shapira: Capturing new developments in an emerging technology: An updated search strategy for identifying nanotechnology research outputs, Scientometrics **95**(1), 351–370 (2013)

39.56 R.N. Kostoff, G.R. Koytcheff, C.G.Y. Lau: Global nanotechnology research literature overview, Technol. Forecast. Soc. Change **74**(9), 1733–1747 (2007)

39.57 Y. Zhang, X. Zhou, A.L. Porter, J.M.V. Gomila, A. Yan: Triple Helix innovation in China's dye-sensitized solar cell industry: Hybrid methods with semantic TRIZ and technology roadmapping, Scientometrics **99**(1), 55–75 (2014)

39.58 R.N. Kostoff: Literature-related discovery: Potential treatments and preventatives for SARS, Technol. Forecast. Soc. Change **78**(7), 1164–1173 (2011)

39.59 R.N. Kostoff, J.L. Solka, R.L. Rushenberg, J.A. Wyatt: Literature-related discovery (LRD): Water purification, Technol. Forecast. Soc. Change **75**(2), 256–275 (2008)

39.60 R.N. Kostoff: Literature-related discovery: Common factors for Parkinson's disease and Crohn's disease, Scientometrics **100**(3), 623–657 (2014)

39.61 N.R. Smalheiser: Literature-based discovery: Beyond the ABCs, J. Am. Soc. Inf. Sci. Technol. **63**(2), 218–224 (2012)

39.62 Y.-N. Tu, J.-L. Seng: Indices of novelty for emerging topic detection, Inf. Process. Manag. **48**(2), 303–325 (2012)

39.63 W. Tang, F.S. Tsai, L. Chen: Blended metrics for novel sentence mining, Expert Syst. Appl. **37**(7), 5172–5177 (2010)

39.64 L.C. Chen, T.J. Yu, C.J. Hsieh: Key graph-based chance discovery for exploring the development of e-commerce topics, Scientometrics **95**(1), 257–275 (2013)

39.65 D. Delen, M.D. Crossland: Seeding the survey and analysis of research literature with text mining, Expert Syst. Appl. **34**(3), 1707–1720 (2008)

39.66 P. Yang, W. Gao, Q. Tan, K.-F. Wong: A link-bridged topic model for cross-domain document classification, Inf. Process. Manag. **49**(6), 1181–1193 (2013)

39.67 Y.-H. Tseng, C.-J. Lin, Y.-I. Lin: Text mining techniques for patent analysis, Inf. Process. Manag. **43**(5), 1216–1247 (2007)

39.68 Y.-H. Tseng, Y.-M. Wang, Y.-I. Lin, C.-J. Lin, D.-W. Juang: Patent surrogate extraction and evaluation in the context of patent mapping, J. Inf. Sci. **33**(6), 718–736 (2007)

39.69 A.J.C. Trappey, C.V. Trappey, C.-Y. Wu: Automatic patent document summarization for collaborative knowledge systems and services, J. Syst. Sci. Syst. Eng. **18**(1), 71–94 (2009)

39.70 C. Son, Y. Suh, J. Jeon, Y. Park: Development of a GTM-based patent map for identifying patent vacuums, Expert Syst. Appl. **39**(3), 2489–2500 (2012)

39.71 Y.-R. Li, L.-H. Wang, C.-F. Hong: Extracting the significant-rare keywords for patent analysis, Expert Syst. Appl. **36**(3), 5200–5204 (2009)

39.72 H. Park, J. Yoon, K. Kim: Identification and evaluation of corporations for merger and acquisition strategies using patent information and text mining, Scientometrics **97**(3), 883–909 (2013)

39.73 J. Choi, Y.-S. Hwang: Patent keyword network analysis for improving technology development efficiency, Technol. Forecast. Soc. Change **83**, 170–182 (2014)

39.74 A. Pilkington, R. Dyerson, O. Tissier: The electric vehicle: Patent data as indicators of technological development, World Pat. Inf. **24**(1), 5–12 (2002)

39.75 G.F. Nemet: Inter-technology knowledge spillovers for energy technologies, Energy Econ **34**(5), 1259–1270 (2012)

39.76 A. Moeller, M.G. Moehrle: Completing keyword patent search with semantic patent search: Introducing a semiautomatic iterative method for patent near search based on semantic similarities, Scientometrics **102**(1), 77–96 (2015)

39.77 S.-H. Liu, H.-L. Liao, S.-M. Pi, J.-W. Hu: Development of a patent retrieval and analysis platform: A hybrid approach, Expert Syst. Appl. **38**(6), 7864–7868 (2011)

39.78 S. Venugopalan, V. Rai: Topic based classification and pattern identification in patents, Technol. Forecast. Soc. Change **94**, 236–250 (2015)

39.79 F. Janssens, L. Zhang, B. De Moor, W. Glänzel: Hybrid clustering for validation and improvement of subject-classification schemes, Inf. Process. Manag. **45**(6), 683–702 (2009)

39.80 X. Liu, S. Yu, F. Janssens, W. Glänzel, Y. Moreau, B. De Moor: Weighted hybrid clustering by combining text mining and bibliometrics on a large-scale journal database, J. Assoc. Inf. Sci. Technol. **61**(6), 1105–1119 (2010)

39.81 B. Yoon, S. Lee, G. Lee: Development and application of a keyword-based knowledge map for effective R&D planning, Scientometrics **85**(3), 803–820 (2010)

39.82 T. Barnickel, J. Weston, R. Collobert, H.-W. Mewes, V. Stümpflen: Large scale application of neural network based semantic role labeling for automated relation extraction from biomedical texts, PLoS ONE **4**(7), e6393 (2009)

39.83 A.L. Porter, I. Rafols: Is science becoming more interdisciplinary? Measuring and mapping six re-

search fields over time, Scientometrics **81**(3), 719–745 (2009)
39.84 National Academy of Engineering, National Academy of Sciences, Institute of Medicine: *Facilitating Interdisciplinary Research* (National Academies Press, Washington, D.C. 2004)
39.85 L.G. Nichols: A topic model approach to measuring interdisciplinarity at the national science foundation, Scientometrics **100**(3), 741–754 (2014)
39.86 A.J. Toynbee: Introduction: The geneses of civilisations. In: *A Study of History*, ed. by A.J. Toynbee (Oxford Univ. Press, Oxford 1963)
39.87 A. Verbeek, K. Debackere, M. Luwel: Linking science to technology: Using bibliographic references in patents to build linkage schemes, Scientometrics **54**(3), 399–420 (2002)
39.88 F. Narin, D. Olivastro: Status report: Linkage between technology and science, Res. Policy **21**(3), 237–249 (1992)
39.89 M. Meyer: Tracing knowledge flows in innovation systems—An informetric perspective on future research science-based innovation, Econ. Syst. Res. **14**(4), 323–344 (2002)
39.90 T. Magerman, B. Van Looy, X. Song: Exploring the feasibility and accuracy of Latent Semantic Analysis based text mining techniques to detect similarity between patent documents and scientific publications, Scientometrics **82**(2), 289–306 (2010)
39.91 T. Magerman, B. Van Looy, K. Debackere: Does involvement in patenting jeopardize ones academic footprint? An analysis of patent-paper pairs in biotechnology, Res. Policy **44**(9), 1702–1713 (2015)
39.92 X. Wang, H. Tian, M.A. Mohammad, C. Li, C. Wu, Y. Yang, T.-L. Ren: A spectrally tunable all-graphene-based flexible field-effect light-emitting device, Nat. Commun. **6**, 7767 (2015)
39.93 D. Thorleuchter, D. Van den Poel, A. Prinzie: Mining ideas from textual information, Expert Syst. Appl. **37**(10), 7182–7188 (2010)
39.94 B. Pang, L. Lee: Opinion mining and sentiment analysis, Found. Trends Inf. Retri. **2**(12), 1–135 (2008)
39.95 A. Pons-Porrata, R. Berlanga-Llavori, J. Ruiz-Shulcloper: Topic discovery based on text mining techniques, Inf. Process. Manag. **43**(3), 752–768 (2007)
39.96 Z. Ma, A. Sun, G. Cong: On predicting the popularity of newly emerging hashtags in Twitter, J. Am. Soc. Inf. Sci. Technol. **64**(7), 1399–1410 (2013)
39.97 Y. Lu, P. Zhang, J. Liu, J. Li, S. Deng: Health-related hot topic detection in online communities using text clustering, PLoS ONE **8**(2), e56221 (2013)
39.98 R. Feldman, J. Sanger: *Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data* (Cambridge Univ. Press, Cambridge 2006)
39.99 Y. Zhang, A.L. Porter, Z. Hu, Y. Guo, N.C. Newman: Term clumping for technical intelligence: A case study on dye-sensitized solar cells, Technol. Forecast. Soc. Change **85**, 26–39 (2014)
39.100 J. Duan, M. Zhang, W. Jingzhong, Y. Xu: A hybrid framework to extract bilingual multiword expression from free text, Expert Syst. Appl. **38**(1), 314–320 (2011)
39.101 T. Wei, Y. Lu, H. Chang, Q. Zhou, X. Bao: A semantic approach for text clustering using WordNet and lexical chains, Expert Syst. Appl. **42**(4), 2264–2275 (2015)
39.102 N. Li, D.D. Wu: Using text mining and sentiment analysis for online forums hotspot detection and forecast, Decis. Support Syst. **48**(2), 354–368 (2010)
39.103 W. He, S. Zha, L. Li: Social media competitive analysis and text mining: A case study in the pizza industry, Int. J. Inf. Manag. **33**(3), 464–472 (2013)
39.104 E.J. de Fortuny, T. De Smedt, D. Martens, W. Daelemans: Media coverage in times of political crisis: A text mining approach, Expert Syst. Appl. **39**(14), 11616–11622 (2012)
39.105 G. Salton, A. Wong, C.S. Yang: A vector space model for automatic indexing, Commun. ACM **18**(11), 613–620 (1975)
39.106 S.-S. Lin: A document classification and retrieval system for R&D in semiconductor industry? A hybrid approach, Expert Syst. Appl. **36**(3), 4753–4764 (2009)
39.107 Y.-L. Chen, Y.-T. Chiu: Vector space model for patent documents with hierarchical class labels, J. Inf. Sci. **38**(3), 222–233 (2012)
39.108 W. Zhang, T. Yoshida, X. Tang: A comparative study of TF*IDF, LSI and multi-words for text classification, Expert Syst. Appl. **38**(3), 2758–2765 (2011)
39.109 K.W. Boyack, D. Newman, R.J. Duhon, R. Klavans, M. Patek, J.R. Biberstine, B. Schijvenaars, A. Skupin, N. Ma, K. Börner: Clustering more than two million biomedical publications: Comparing the accuracies of nine text-based similarity approaches, PLoS ONE **6**(3), e18029 (2011)
39.110 Z.H. Deng, K.H. Luo, H.L. Yu: A study of supervised term weighting scheme for sentiment analysis, Expert Syst. Appl. **41**(7), 3506–3513 (2014)
39.111 Z. Hu, S. Fang, T. Liang: Empirical study of constructing a knowledge organization system of patent documents using topic modeling, Scientometrics **100**(3), 787–795 (2014)
39.112 J. Ma, A.L. Porter: Analyzing patent topical information to identify technology pathways and potential opportunities, Scientometrics **102**(1), 811–827 (2015)
39.113 K.K. Bharti, P.K. Singh: Hybrid dimension reduction by integrating feature selection with feature extraction method for text clustering, Expert Syst. Appl. **42**(6), 3105–3114 (2015)
39.114 S. Choi, H. Park, D. Kang, J.Y. Lee, K. Kim: An SAO-based text mining approach to building a technology tree for technology planning, Expert Syst. Appl. **39**(13), 11443–11455 (2012)
39.115 Y. Li, J. Shawe-Taylor: Advanced learning algorithms for cross-language patent retrieval and classification, Inf. Process. Manag. **43**(5), 1183–1199 (2007)
39.116 P. Burnap, O.F. Rana, N. Avis, M. Williams, W. Housley, A. Edwards, J. Morgan, L. Sloan: Detecting tension in online communities with

computational Twitter analysis, Technol. Forecast. Soc. Change **95**, 96–108 (2015)
39.117 C.-P. Wei, C.C. Yang, C.-M. Lin: A latent semantic indexing-based approach to multilingual document clustering, Decis. Support Syst. **45**(3), 606–620 (2008)
39.118 T. Hofmann: Probabilistic latent semantic indexing. In: *Proc. 22nd Annu. Int. ACM SIGIR Conf. Res. Dev. Inf. Retri* (1999) pp. 50–57
39.119 D.M. Blei, A.Y. Ng, M.I. Jordan, J. Lafferty: Latent Dirichlet allocation, J. Mach. Learn. Res. **3**, 2003 (2003)
39.120 M. Song, S.Y. Kim: Detecting the knowledge structure of bioinformatics by mining full-text collections, Scientometrics **96**(1), 183–201 (2013)
39.121 M. Ghiassi, M. Olschimke, B. Moon, P. Arnaudo: Automated text classification using a dynamic artificial neural network model, Expert Syst. Appl. **39**(12), 10967–10976 (2012)
39.122 G.A. Miller, W.G. Charles: Contextual correlates of semantic similarity, Lang. Cogn. Process. **6**(1), 1–28 (1991)
39.123 M. Baroni, G. Dinu, G. Kruszewski: Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In: *Proc. 52nd Annu. Meet. Assoc. Comput. Linguist.* (2014) pp. 238–247
39.124 T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean: Distributed representations of words and phrases and their compositionality, Proc. Adv. Neural Inf. Process. Syst. **26**, 3111–3119 (2013)
39.125 D.M. Blei, J.D. Lafferty: Topic models. In: *Text Mining: Theory and Applications*, ed. by A. Srivastava, M. Sahami (Taylor Francis, New York 2009) pp. 71–93
39.126 A. Suominen, H. Toivanen: Map of science with topic modeling: Comparison of unsupervised learning and human-assigned subject classification, J. Assoc. Inf. Sci. Technol. **67**(10), 2464–2476 (2016)
39.127 F. Rizzi, E. Annunziata, G. Liberati, M. Frey: Technological trajectories in the automotive industry: Are hydrogen technologies still a possibility?, J. Clean. Prod. **66**, 328–336 (2014)
39.128 K. Frenken, M. Hekkert, P. Godfroij: R&D portfolios in environmentally friendly automotive propulsion: Variety, competition and policy implications, Technol. Forecast. Soc. Change **71**(5), 485–507 (2004)
39.129 J.H. Wesseling, J. Faber, M.P. Hekkert: How competitive forces sustain electric vehicle development, Technol. Forecast. Soc. Change **81**, 154–164 (2014)
39.130 P. Kurzweil: History of fuel cells. In: *Encyclopedia of Electrochemical Power Sources*, ed. by J. Garche (Elsevier, Amsterdam 2009)
39.131 M. Cropper, S. Geiger, D. Jollie: Fuel cells: A survey of current developments, J. Power Sources **131**(1/2), 57–61 (2004)
39.132 C.-K. Yau, A. Porter, N. Newman, A. Suominen: Clustering scientific documents with topic modeling, Scientometrics **100**(3), 767–786 (2014)
39.133 A. Suominen, H. Toivanen, M. Seppnen: Firms' knowledge profiles: Mapping patent data with unsupervised learning, Technol. Forecast. Soc. Change **115**, 131–142 (2016)
39.134 W.S. Lee, E.J. Han, S.Y. Sohn: Predicting the pattern of technology convergence using big-data technology on large-scale triadic patents, Technol. Forecast. Soc. Change **100**, 317–329 (2015)
39.135 V.D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre: Fast unfolding of communities in large networks, J. Stat. Mech. **2008**(10), P10008 (2008)
39.136 J. Chang, S. Gerrish, C. Wang, D.M. Blei: Reading tea leaves: How humans interpret topic models, Adv. Neural Inf. Process. Syst. **22**, 288–296 (2009)
39.137 R. Arun, V. Suresh, C.E. Veni Madhavan, M. Narasimha Murty: On finding the natural number of topics with latent dirichlet allocation: Some observations, Lect. Notes Comput. Sci. **6118**, 391–402 (2010)
39.138 A. Suominen, M. Seppänen: Bibliometric data and actual development in technology life cycles: Flaws in assumptions, Foresight **16**(1), 37–53 (2014)

Samira Ranaei
School of Engineering Science, Industrial Engineering and Management
Lappeenranta University of Technology (LUT)
Lappeenranta, Finland
samiratzn@gmail.com

Samira Ranaei is a Doctoral student at the School of Engineering Science at Lappeenranta University of Technology (LUT) in Finland. She has an Industrial Engineering Bachelor degree and received her Master's in Innovation Management from LUT in 2013. Her research interests are Innovation and Technology Management, Technological Forecasting, and Patent Analysis techniques.

Arho Suominen
VTT Technical Research Centre of Finland
Espoo, Finland
arho.suominen@vtt.fi

Arho Suominen is a Senior Scientist at the Innovations, Economy, and Policy Unit of the VTT Technical Research Center of Finland, while also lecturing at the Department of Information Technology at the University of Turku. His research focuses on qualitative and quantitative assessment of innovations systems with a special focus on data mining methods. He has a Doctor of Science (Tech.) degree from the University of Turku.

Alan Porter

Search Technology, Inc.
Norcross, GA, USA
aporter@searchtech.com

Alan Porter is Director of R&D for Search Technology, Inc., Norcross, GA, USA. He is also Professor Emeritus of Industrial & Systems Engineering, and of Public Policy, at Georgia Tech, where he remains Co-director of the Technology Policy and Assessment Center. He received a PhD in Psychology from UCLA in 1972. He is the author of some 240 articles and books. Current research emphasizes measuring, mapping, and forecasting emerging technologies.

Tuomo Kässi

School of Engineering Science, Industrial Engineering and Management
Lappeenranta University of Technology (LUT)
Lappeenranta, Finland
tuomo.kassi@lut.fi

Tuomo Kässi is Professor Emeritus of Industrial Engineering and Management at Lappeenranta University of Technology, Finland. His educational background is in Technical Physics and Economics. He has worked as an engineer and manager in industry. His teaching and research interests are industrial transformation, management of technology, product platforms, and recently IPR issues and patents. Kässi has been a responsible director in many research projects financed by public or public-private partnerships.

40. Functional Patent Classification

Andrea Bonaccorsi, Gualtiero Fantoni, Riccardo Apreda, Donata Gabelloni

Patent classifications are systematically used in patent analysis for a number of purposes. Existing classifications not only shape the administrative activities of recording and reporting and the search for prior art, but also create the backbone of the construction of science and technology indicators used in economic analysis, policy making, and business and competitive intelligence.

Yet the current classification system of patents, despite significant and continuous efforts to update, suffers from a number of limitations. In particular, it fails to capture the full potential of inventions to cut across industrial boundaries, does not allow fine-grained technology intelligence, and misses almost entirely the opportunities for lateral vision.

We suggest integrating existing schemes with a full scale functional classification, i. e., based on the main functions performed by a technology, rather than on the inventive solutions or their potential applications. The functional approach allows us to overcome most of the limits of traditional classification, due to the generality and abstraction of the representation of functions. In this chapter, we will first review the conceptual background of the functional approach in epistemology and analytical philosophy and illustrate its recent developments in engineering design, design theory, artificial intelligence, computational linguistics, and data mining. We then discuss three short case studies of the application of the methodology for the definition of patent sets (in particular within a technology foresight exercise), prior art analysis, and technology crossover identification and mapping.

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_40

40.1 Patent Classifications

Patents are routinely classified into classes by patent offices. The classification process follows international standards, which were established back in history and are regularly updated in order to follow the technological evolution. Classifications are, then, routinely adopted for a large variety of practical and analytical purposes, ranging from prior art search to econometric modeling.

Patent classifications follow a mix of functional and industrial criteria. By *functional* we mean a classification that is based on the engineering principles underlying the technical invention.

In the engineering literature, a function is defined as follows [40.1, p. 72]:

> The function is a property of the technical system, and describes its ability to fulfill a purpose, namely to convert an input measure into a required output under precisely given condition.

A function is a transformation that takes place in the physical space. From a representational point of view, it can be described by a verbal expression (functional verb), usually associated to an object and a modal expression.

This notion is, needless to say, crucial for the working of the entire building of the patent system. Functional descriptions deal with the *novelty* and *utility* requisites of patents. The notion of utility requires that the invention deliver something of value for a user, at least potentially. For the utility to be delivered, what is needed is precisely that a function (perhaps a complex one) is implemented. The notion of novelty requires that this function has never been implemented the way it is implemented in the patent application. More rarely, the function itself is an invention, in the sense that it was not conceptualized before the invention. Industrial criteria, on the contrary, refer to potential users of inventions. They deal mainly with the *industrial* requisite of patents.

The functional language is crucial in the definition of invention, as well as in the doctrine of functional equivalence, stating that solutions that realize the same function should be considered protected by the same patent. Furthermore, it can be shown that some of the most relevant changes in legislation and practice in the last decades, namely the patentability of software and of biotechnological inventions, have been promulgated on the basis of scientific theories about the alleged functional impact of patentability [40.2].

Notwithstanding the foundational role of the concept, legal and economic doctrines and practices rely on a relatively informal definition of functions. The International Patent Classification (IPC) and the Cooperative Patent Classification (CPC), for example, use a relatively loose definition of functions, and use it in a non-systematic and complete way. Furthermore, there has not been any systematic connection with the disciplines in which a great deal of work has been done to define and formalize functions, such as theories of engineering design, theories of systematic invention, or functional analysis.

In this chapter, we argue that the current classification system of patents, despite a significant effort to update, is not able to follow the acceleration of technological developments. New pervasive and transversal technologies, broadly defined as digital technologies, have almost destroyed industrial boundaries and opened new forms of lateral and transversal competition. New products and services are developed by combining existing technologies in novel ways. Very often inventions that were initially conceived for an industry find applications in entirely new domains. To mention just a couple of examples, blockchain solutions, which were initially conceived for the financial industry, now find applications in agriculture, in order to certify the space and time of an inspection of a protected label. Or Second Life, which was intended as an entertainment software platform, is currently largely used as a therapeutic tool.

Under this kind of technological dynamics, existing classifications are almost inevitably in delay. In particular, mixed functional-industrial classifications fail to capture the full potential of inventions to cut across industrial boundaries. Consequently, using existing patent classes does not allow a fine-grained technology intelligence and misses almost entirely the opportunities for lateral vision.

We suggest integrating the existing classifications with a full scale functional classification, based on functional hierarchies and supported for specific tasks by the construction of a large functional dictionary.

The chapter is structured as follows. First, we offer a short overview of the notion of function in a variety of disciplines and comment on recent advancements in computational linguistics that have made it possible to develop large scale dictionaries and classifications. In Sect. 40.3, we discuss in detail the limitations of existing patent classifications. In Sect. 40.4, we offer three short case studies of the application of the methodology. The final section concludes this chapter.

40.2 A Brief History of Functional Analysis

The main elements that characterize functions are as follows:

1. Functions are *abstract* representations–they must be independent on specific technical solutions. This is called solution neutrality.
2. Functions are *normative*–they describe a purpose, or a goal, or a raison d'être of an object, and in this way they describe the conditions under which the object may come into existence, and, correspondingly, the conditions under which the object may not work properly, or its dysfunction.
3. Functions are *hierarchical*–they can be decomposed in an iterative way, moving up to functions of higher abstraction or down to functions of lower abstraction, or higher instantiation. However, there might be multiple hierarchies or different ways to decompose a higher level function.
4. Functions involve a *transformation in the physical space*–they can then be made consistent with physical descriptions.

Due to the novelty of the methodology, it is useful to review briefly the origins of the main concepts and definitions. The notion of function is an established one in engineering and architecture disciplines and has attracted the attention of philosophers, but it has not yet gained universal recognition due to a number of theoretical and practical issues, which we must illustrate in detail.

In the following, we call the attention of the community working on Science and Technology (S&T) indicators to the deeper intellectual roots of some of the concepts that we utilize.

40.2.1 Philosophical Foundations of Functional Thinking

According to Aristotle, the general notion of cause, as conceptualized for physical entities, was not sufficient to explain the relation between some forms of action and the world. In his *Physics*, Aristotle introduced the notion of *telos*, or goal, as the basis for a separate form of explanation, called teleological explanation [40.3, p. 8]:

> Teleological explanation in Aristotle pertains broadly to goal-directed actions or behavior. Aristotle invokes teleology when an event or action pertains to goals: 'that for the sake of which'.

The explanation was found by positing a separate notion of cause, called *final cause*. Final causality worked backward: the existence of a final goal required by necessity the working of individual elements in such a way that their coordination could ensure the working of the whole organism. Interestingly, Aristotle used the notion of *telos* to cover two distinct kinds of final teleology: agency-centered teleology (involving behavior and artifacts, or technology) and teleology pertaining to natural organisms [40.3]. While the distinction between the two classes is one of consciousness (agents are aware of the goal of their behavior, natural organisms are not), the notion of functions can be applied to both fields. According to Aristotle, consequently, the actions aimed at constructing artifacts have value only as manifestations of human goals, that is, they have no internal necessity [40.4].

This notion was rejected after the modern Scientific Revolution. Galileo already asked to reason only in terms of physical causes, leaving the overall *why* to things outside the scientific domain. Notably Darwin, in his rejection of Lamarck's explanation of the adaptation of life, called for a scientific reasoning that moved only from causes to effects and never backward [40.5, 6]. Adaptation comes from random variations that are selected only by virtue of their fitness to the environment and not by virtue of some finalization to superior forms. Variation is random, selection is blind. This is *function without purpose* [40.7, 8].

These arguments were formalized in the twentieth century in two separate traditions. On the one hand, the neo-Darwinian synthesis, introduced by authors such as S. Wright, R.A. Fisher, and J. Maynard Smith, provided the mathematical framework to examine the way in which random genotypic variations could generate new forms or new phenotypes [40.6, 9–11]. On the other hand, the neopositivist philosophy of science rejected the teleological arguments, proposing that arguments outside the physical causality notion, introducing teleology, should not be considered valid [40.12, 13].

Overall, there was a general consensus until the end of the twentieth century on the idea that the notion of function, having inevitably a normative content, should not be admitted in scientific reasoning at all.

This agreement encountered some trouble at the end of the twentieth century, although the streams of thinkers that have criticized the dominant view are still somewhat peripheral. In Kuhnian terms, they point to anomalies in the dominant view, but they still have not gained central ground. Nevertheless, the convergence of critical arguments from a variety of theoretical starting points is an indication of deeper problems.

First, within the philosophy of biology, the notion of pure randomness of variations has been criticized. Starting from the famous metaphor of the spandrels of San Marco in Venice [40.14], *Gould* argued vigorously that variation is not random, but has an internal structure, most likely hierarchical, that makes the probabilities to change in different directions in the space of genetic possibilities highly constrained [40.15–17]. This argument, initially minoritary in evolutionary biology, found support in the recent advancements of epigenetics and systems biology.

Second, the notion of organism has gained prominence in a recent revival of the philosophy of Hegel. The German philosopher tried to conceptualize the existence of complex living entities by positing a new type of causality, called dialectics, as a mutual determination of constraints at different hierarchical levels of reality. Here, the main question is not whether causality can work backward in time (as in teleology) but can work downward, or from a higher-level structure or organism, down to more elementary units. The extent to which dialectics could be used in the philosophy of science is open to debate, but it is clearly on the table. It seems to be a way to give autonomous philosophical foundations to the notion of organization, which is also central in the analysis of complex systems.

Finally, in the last two decades, a number of contributions in analytical philosophy have reintroduced and rejuvenated the notion of function, in a way fully compatible with the procedures of scientific reasoning. In this tradition, starting from the earlier studies of *Wright* [40.18] and *Cummins* [40.19], biological and technological systems are considered an instantiation of a more general class of entities in which the internal working satisfies some general conditions, drawn from the scientific knowledge of the world, for survival or self-reproduction [40.20–24]. Functions may be described in causal terms, according to this analytic tradition, although with some qualifications [40.25, 26].

Summing up, and to make a long story short (for a longer story see [40.27]), we now witness an intellectual landscape in foundational disciplines that is favourable to the development of a general framework, which gives prominence to the notion of functions. This intellectual climate fits nicely with recent developments in the applied disciplines, such as engineering or design, to which we now turn.

The following is not a complete review of existing approaches to functional analysis, which would require a long exposition. There are several traditions, mainly developed in USA, Germany, Russia, and Japan, with a large number of contributions. Rather, we try to identify those contributions that are more relevant for our issue: how to use functional analysis in the field of patent classification and patent search.

40.2.2 The German School of Systematic Engineering Design

While the notion of function is intuitive, the challenge is how to represent it in such a way to be able to manipulate and use the concept in practical terms. This is the object of functional analysis, a stream of scholarship in engineering design whose goal is to develop theoretical frameworks and tools to represent technical problems in an abstract way.

This challenge was taken up systematically by a number of authors, mainly in mechanical engineering, who wrote in German and were active in German-speaking countries. Their books have subsequently been translated into English. Examples of this approach are *Hubka* and *Eder*'s *Theory of Technical Systems* [40.1] and *Pahl* and *Beitz*'s *Engineering Design. A Systematic Approach* [40.28, first edition 1977].

These fundamental books should be interpreted in the light of the professional tradition of German engineers and of their academic training (for a reconstruction of this tradition see [40.29]). German engineers are trained to design new technical solutions by moving from first principles in a highly structured way. The specific rules of reasoning are written down in a formal way and generate lists of suggestions. In some sense, the discipline is one of the systematic examination of many solutions, following a formal list of methods to guide the reasoning process (see also [40.30]).

It is within this tradition that one should understand functional analysis, as a systematic effort to describe and standardize the abstract requirements of design tasks. A development that took place independently but with a similar abstraction goal (and a rather similar time trajectory), should also be mentioned here, i. e., the theory of inventive principles, or TRIZ (the Russian acronym for the *theory of inventive problem solving* [40.31]). While this is not the object of this chapter, the TRIZ literature pushed forward the notion that there is a compact collection of general inventive principles to be applied to engineering problems, whose functional meaning can be made explicit. Indeed, even if treated from a different point of view, functional thinking has a significant role in TRIZ theory.

40.2.3 Artificial Intelligence and Design: From Herbert Simon to the Carnegie Mellon Project

From an entirely different intellectual tradition, a number of scholars of artificial intelligence started to ask whether the design task could be formalized and automatized. This was in clear continuity of the ambitious tradition of cognitive science that started from the automation of well-structured problems, such as chess playing, and moved to ill-structured problems, such as scientific discovery or design [40.32–35]. At Carnegie Mellon, this project was pursued systematically for many years [40.36–40].

The emphasis here was more on the cognitive procedures utilized to move in the design space, called heuristics, than on the initial requirements of design tasks. Nevertheless, the formal definition of requirements in terms of functions was a necessary element of the description of the problem [40.41]. A remarkable example of this tradition is *Tong* and *Sriram*'s *Artificial Intelligence in Engineering Design* [40.42] and *Sriram*'s *Intelligent Systems for Engineering* [40.43].

Interestingly, these efforts did not produce compelling results. The most relevant applications were found in the field of the design of electric and electronic circuits. Applications in the field of architectural design were also explored, but with limited success.

One might argue that the focus on the cognitive procedures missed a point, that is, the language with which people transform the abstract requirement (the *why* of the artifact) into broad ideas, then more precise concepts, down to detailed specifications and design. In other words, there must be an interaction between cognitive processes of the general type [40.44, 45] and the specific representational language in each of the design domains. Such interaction was almost entirely missed in earlier efforts to automatize design, due to the emphasis on discovering and modeling general mechanisms of intelligent behavior [40.35].

40.2.4 Functional Bases

These limitations led to a different strategy, i. e., focusing more on the language than on the cognitive operations in the design space. The task was to develop a language that might be used to write functional expressions that were consistent with grammar and semantic rules.

Interestingly, this intuition was forwarded in the direction of the development of functional languages based on the manipulation of a very *small* set of functional verbs. This strategy followed some sort of Occam's razor argument: the goal was to describe functionally an artifact using the smallest possible number of different functional verbs. The main research goal was parsimony and elegance, rather than coverage and usability in every context. This approach, labeled *functional basis*, was developed mainly in the USA, starting with the pioneering works of *Little* et al. [40.46] and *Stone* and *Wood* [40.47]. A more systematic version was elaborated shortly after by *Hirtz* et al. [40.48] and further expanded by many authors [40.49–51]. The functional basis paradigm created a systematic linkage with earlier traditions of engineering analysis, such as value analysis. Several US research teams created a stream of research that transformed functional analysis into a usable tool.

Over time, however, this approach showed some limitations; in real world applications, it was practically impossible to use functional bases without the help of experts in specific engineering domains. In order to capture the functions of artifacts it was necessary to add several qualifications to verbs that were too broad or general. These limitations were evident when the trend towards automatic text processing became dominant. Functional bases were mainly intended for manual use. They were not suited for the treatment of large collections of texts.

40.2.5 Introducing Behavior in the Functional Representation: The Function–Behavior–Structure (FBS) Model

The limits of functional bases were clearly anticipated in a stream of literature that introduced a new layer of description.

In functional analysis, a distinction is made between the structure of the artifact (i. e., its geometry and material composition) and the function, or the abstract description of its purpose. The function is, however, implemented by the structure of the artifact in a dynamic way, that is, by producing a behavior. This behavior is consistent with the structure and is aimed at delivering the function. In functional analysis, the description of the behavior is absorbed in the functional description, by working with varying degrees of granularity in the hierarchy of functions.

Various authors [40.52, 53] suggested enriching the framework by developing a separate layer, called behavior, within a unitary framework called *function–behavior–structure* (FBS). In this way, functional descriptions can be left more general, and the implementation of functions can be described more carefully in dynamic terms, by following the behavior of artifacts in their context [40.53–56]. This approach is much more flexible and articulated, as it permits the representation of functions in dynamic terms, as well as a more explicit link between the functions and the expected behavior or user expectation [40.57–60]. In other words, in linking structure and function through the notion of behavior, this approach allows us to examine expected behavior by users as an indication of needs. It has received large acclaim in the literature [40.61, 62].

Other authors have also articulated functional analysis in order to accommodate a separate behavior layer [40.52, 63].

40.2.6 The Ontology Revolution and the Role of Computational Linguistics

The developments discussed above (Pahl and Beitz's systematic design, functional bases, and the FBS framework) were developed within the engineering design literature. This literature is produced by a relatively small community, whose main interests are in the construction of formal systems for representing technological problems moving from a deep knowledge of engineering disciplines [40.64]. In other words, this is a minority of scholars in engineering who, in addition to, sometimes (more rarely) in substitution for, deep studies in specialized engineering disciplines, have a propensity for theoretical generalization and formalization.

Parallel to these developments, and initially with no overlappings, the last two decades have witnessed impressive advancements in computational linguistics, and the ability of artificial systems to process large collections of texts has increased enormously. These developments are based on the construction of formal ontologies or abstract representations of entities and their relations. In parallel, powerful statistical methods for the extraction of meaningful information have been developed in the fields of information retrieval and data mining.

After several pioneering contributions, a full-scale effort to develop a functional ontology was promoted by *Kitamura* and co-authors [40.65–68]. The construction of an ontology requires the formal modeling, using knowledge representation concepts and theorems, of the substantive relations among entities [40.69]. This is usually done in interaction between domain knowledge experts (in this case, engineers and designers) and computer science experts.

After the initial effort, the literature on functional ontologies witnessed large adoption [40.70]. Yet, other authors followed a different path: they built functional representations by massively processing technical texts (in particular, patents) in order to automatically extract functional information. This was done without a pre-existing functional ontology, but only on linguistic bases. In particular, *Montecchi* and *Russo* started to apply to the newly created patent classification (CPC) linguistic queries based on the FBS framework and its variants [40.71–73].

Other authors did not capitalize on existing engineering functional frameworks but reconstructed the notion of function on the basis of purely linguistic structures, e.g., actions. This approach is called SAO. The acronym SAO means subject-action-object structures: every verbal construct in which there is a subject doing an action that involves an object. *Yoon* and *Kim* [40.74] developed a patent analysis strategy based on the concepts of natural language processing. Taking advantage of parsing tools like the Stanford Parser or Knowledgist, patents claims are analyzed to reach a level where each of them is representable by a set of SAO structures. This permits a rapid identification of what the components in a new product are, and what their function is. The extracted structures are compared using a similarity measure that is purely linguistic, i.e., does not implement an underlying ontology. A similar approach is found in *Choi* et al. [40.75], while *Park* et al. [40.76] combine the SAO structure with TRIZ inventive principles. These contributions realize the goal of offering a rich language for functional descriptions. They might be labeled *text mining without ontologies*. The underlying engineering, and ultimately physical, constitution is not made manifest.

40.2.7 Functional Dictionaries

Finally, an alternative path was followed at the intersection between engineering design and machine learning, i.e., the development of large *functional dictionaries*. This path followed the notion that cognitive operations in design depend on the representation of the design task in specific, contextual, and semantically rich environments, in which the interplay between function and structure could be produced. This idea was also at the origin of the Functional Basis movement of Hirtz, Stone, Wood, and co-authors. These authors, however, pursued a goal of parsimony.

The functional dictionary approach goes in the opposite direction, developing the *largest* possible dictionary. It also goes in a different direction to the ontology movement of Kitamura and co-authors, in the sense that there is no need to develop a full scale ontology. What is needed is a procedure to generate the largest possible collection of functional lemmas and to demonstrate that the semantic space of design is *saturated*, or there are not any important undescribed elements left. By saturation, we mean a methodology, borrowed from social sciences, in which there is systematic interaction between observed data and modeling until the model includes all elements that are needed to explain the data. In the context of engineering design, this means completeness at two levels: the level of the *categories* of elements needed to describe an artifact and the level of all the various possible *technical implementations* existing in each category. A related crucial factor, to be dealt with properly for the methodology to work, is the domain dependency of many types of technical concepts: saturation in one domain does not guarantee the same in another, and completeness must be achieved at a third level too, that of all sectors of interest. In our case, we tested the dictionary in a dozen applications in highly disparate industries, up to the point where the entries were fully adequate to describe the problem at hand. The construction of large repositories of technical terminology has been investigated by various research groups, especially within the TRIZ community. As an example, the commercial software Goldfire Innovator (http://inventionmachine.com) includes a large database of functions and physical effects; furthermore, the extraction of functional information from patents was pioneered by *Cascini* and co-authors [40.77–79]. Dictionaries for specific sectors have been built by a number of authors (see, for example, [40.80, 81]).

This direction was taken over a decade ago also by the authors of this chapter [40.27, 82–86]. To reach the optimal result, we combined different techniques, ranging from advanced text mining, use of knowledge patterns and structures, and human factor analysis, all revised by experts in the technological sectors examined.

The most important achievement of this research effort was the construction of a functional dictionary containing more than 100 000 lemmas, of which there are approximately 12 000 functional verbs. This dictionary contains functions, behaviors, and structures, defined as atoms of the artifact, in terms of process, action, or task that the artifact system is able to perform. All entries are related to semantically related entries, such as synonyms, antonyms, and hyperonyms.

This approach has opened the way to an automatic procedure to extract functional information from patents [40.87], while keeping full control of the underlying physical description of functions. This dictionary has been repeatedly used in tasks of patent search, patent classification, topics modeling, technology foresight, and design crossover, in the last few years.

More recently, the same dictionary approach was followed in an effort to build up other technical dictionaries referring to advantages/disadvantages of artifacts and to users or stakeholders of artifacts. The idea is to increase the coverage of all the directions in the design space, thus expanding the possible applications. Overall, the following dictionaries have been developed:

- Stakeholders: persons who have relations with a product or service. It has been built by merging multiple lists (e. g., users, workers, patients). Its size is about 77 000 entries [40.88].
- Advantages and disadvantages: positive and negative effects of products/services. These classes could be also defined as benefits and failures. They consist of more than 20 000 entries.
- Components: list of systems and sub-systems contained in products.
- Physical quantities and units of measurement: physical properties of a phenomenon that can be quantified by numbers and units of measurement.

More recently, a *vertical* dictionary has been built, called Technimeter® 4.0 [40.89]. It is a list of technologies and techniques related to Industry 4.0. The dictionary and taxonomy behind the Technimeter are designed to map documents of the new industrial revolution. It has the form of a fully linked graph and consists of about 2000 technologies and 200 000 links. It is in three languages (automatically expandable) and could be easily extended to fields like precision agriculture, the Internet of Things (IoT), smart cities, smart energy, and E-health.

A sample of entries from these dictionaries is given in Table 40.1.

The approach, which is based on large, non-ontological dictionaries is not suited for all type of analyses, at least in its current state of development. When the analysis requires a high level of abstraction with respect to the specific artifact descriptions, such as in the construction of functional diagrams or other functional modeling-related tasks, and in general for every study performed by a human expert, the variety and the details embedded in the database just add unneeded complexity. In such cases, more prescriptive and concise methods, such as those built on functional bases, usually deliver faster results.

On the other hand, in all cases in which the investigation needs to deal with the complexity and fuzziness of natural language, a complete functional dictionary provides, almost by design, an efficient and reliable instrument. This is the case of all forms of software-based, automated analyses of texts, which in turn are the only possible way of tackling large amounts of technical documents in tasks as diverse as information retrieval, knowledge extraction, or document categorization and labeling.

One of the most promising areas of application of the functional dictionary is, indeed, patent classification. We suggest that a full scale functional dictionary allows a fine-grained representation (i. e., retrieval, mapping, clustering, and profiling) of patent information, opening the way to a variety of powerful applications. As a future development, integrating a full dictionary with computational linguistic algorithms may even allow tasks such as constructing functional diagrams in an automated way and comparing them.

In the next chapters, alongside the general discussion about functional classification of patents, we will also provide some examples of use, among others, of the dictionary approach.

Table 40.1 Sample of entries from the functional dictionary, the stakeholder or user dictionary, the advantage/disadvantage dictionary and the Technimeter 4.0 (in alphabetical order)

Functional behavior structure			Stakeholders, pains and gains			Technological aspects	
Functional verbs	Behaviors	Structures/ components	Users	Advantages	Dis-advantages	Units of measurement	Technimeter® 4.0
Abrade	Abraham–Lorentz (force)	Annular, annulus	Academician	Ability	Abandon	ag	Actroid
Absorb	Absorption	Axial, axis	Accordionist	Accessible	Absent	aHz	ADA (programming language)
Abut	Acoustic (shock, wave, ...)	Arc, arched, arcuate	Acoustician	Accommodate	Abuse	aJ	AM (additive manufacturing)
Accelerate	Activation (energy, coefficient)	Ampoule	Acrobat	Accurate	Accident	am/s^2	Advanced audio distribution profile
Access	Adhesion, adhesive	Aperture	Actor	Accurateness	Accidental	aN	Advanced mobile phone system
Acidify	Adiabatic (transformation)	Arm	Acupuncturist	Adaptable	Aggravated	A	AGV (automated guided vehicle)
Acierate	Archimedes (principle)	Axle	Adult	Adequate	Aggravation	Ampere	Aibo (artificial intelligence robot)
Adapt	Auger (effect)	Anode	Aesthetician	Adjustability	Alterations	atto-	AIML (artificial intelligence makeup language)
Add	Austenitization	Antenna	Alpinist	Adjustable	Anxiety	aW	Air cobot
...	...	...	...	...	...	...	...

40.3 Patent Search and the Limitations of Existing Patent Classifications

Patent classification is a necessary part of any patent system, for legal, administrative, and practical purposes. One of the main areas of utilization of patent classification is patent search, which has two main applications: ex ante patent search or the search of prior art done by inventors, assignees, attorneys, and patent officers before and during patent application, and ex post patent search, or the search in databases carried out after the publication of patents.

Patent search can be based on several alternative strategies of query. Some queries are exclusively based on existing patent classifications (IPC or CPC) or industrial classifications, others use other metadata that are included in patent documents. Summing up a patent search is usually done in one or more of the following ways:

1. IPC or CPC classifications
2. The codes of NACE-CLIO, that is, the European Industry Classification (The acronym stands for Nomenclature générale des Activités économiques dans les Communautés Européennes–Classification Input–Output or equivalent national correspondents, such as the Italian ATECO (ATtività ECOnomiche), to identify the name of companies of the field of interest
3. Keywords associated with technologies of interest
4. Full names of companies and/or research centers that develop technologies in the field of interest
5. Full names of inventors.

Interestingly, each of these search strategies suffers from a number of severe limitations. We review them in order.

40.3.1 IPC or CPC Classes

The official patent classifications are the most largely used in patent analysis, both in professional practice and in academic research. In the latter domain, patent classifications are routinely used in the economics of innovation and strategic management literature, in order to address issues such as diversification of companies, related variety, innovation search, or novelty. Yet these classifications suffer from limitations that are not always clearly recognized.

To start with, the large number of IPC codes (more than 70 000 IPC codes among classes, subclasses, groups, and subgroups [40.90] and of CPC entries (more than 200 000) [40.91], while an indication of the effort of IPR authorities to follow the evolution of technology, generates a cumbersome task. Patent officers and analysts are faced with a severe trade-off: using fine-grained classification requires a large specialized knowledge, while using higher-level codes would bring in the patent set lot of noise from distant and unrelated documents. As a matter of fact, the reading of the definitions of patent classes does not solve at all uncertainties in classification and also in information retrieval.

Second, the IPC/CPC classification has its own ambiguities of attribution. Compare, for example, the subgroup *A61B 5/00–Measuring for Diagnostic Purposes* with the class *G01 Measuring*, or the subclass *F16F–Springs; Shock Absorbers* with the equally valid subclass *B60G–Vehicle Suspensions*. It is clear that a patent of interest can be legitimately listed under one or the other code. This means that an incomplete IPC-based query (in particular, a query that fails to recognize these kinds of ambiguities) will miss important information.

Third, there are errors of classification. In some cases, the technology covered by the patent has nothing to do with the patent class, due to mistakes or misprints. However, the most intriguing (and disturbing) classification error is intentional. Applicants submit patent applications that intentionally include misleading information, that lead patent analysts at patent office into misclassification. In other words, companies try to hide the true content of their patents from competitors, for defensive purposes or for creating hidden threats. It is not uncommon, for example, to see inventions for power windows in the automotive sector classified as blinds for use in houses or solutions for gas turbines classified as solutions for standard combustion engines, and vice versa.

Fourth, the speed at which new patent classifications are introduced does not match the speed of technological evolution. Despite significant efforts, official classifications are several steps behind the technological state of the art. In particular, patent classifications are under pressure in following inventions that are transversal in nature. In general, patents with broad cross-field and cross-industry application are classified in several sectors of the IPC classification. On the other hand, CPC has tried to address the issue by introducing the Y class, but for the time being the coverage is far from complete. As a clear example, it has been shown [40.92] that only a tiny fraction (5%) of the relevant patents in the field of bioinformatics is listed under the corresponding IPC code G06F19/10, while all the others are scattered among over 30 codes. A similar problem refers to the case of transversal or interdisciplinary technologies, which adopt several technical solutions and span several applications, and, therefore, can be pertinent to many classes. Consider for example patents for robotics, or IoT. More generally, even standard technologies may have multiple IPC/CPC attributions. A classic case is control software, which can be classified either under pure software classes or under classes related to the specific industrial sector of application.

Finally, the IPC/CPC classification has not yet been adopted in all National Patent Offices.

Summing up, the use of IPC or CPC classification schemes is justified as a first approximation, while it suffers from severe limitations if the goal is to identify emerging technologies and technological trends, as well as to build up strategic technology intelligence tools that allow for lateral innovation, boundary-crossing technologies, and strategic hiding behavior by competitors.

40.3.2 Industry Codes

A suitable alternative to the construction of lists of companies is to rely on industry codes. At the European level, they follow the NACE-CLIO nomenclature.

This approach, too, has various limitations. First, using industry classification creates the same problem of classification errors found for patent classes; sometimes companies are listed in classes that have nothing to do with the reality of their production, due to misclassification.

Secondly, large enterprises and holdings generally operate in more than one industrial sector; thus the reference is to multiple NACE codes, so that a clear association of enterprises to NACE is difficult. Furthermore, in the case of groups or holdings, quite often the parent company is classified under services, although the associated or subsidiary companies are manufacturing enterprises.

Finally, research centers are not classified by industry codes.

40.3.3 Keywords

Keywords are another largely used technique in patent search. After patent classes, keywords are probably the most important search tool. Keywords must be built up after an expert judgment. More recently, the elicitation of keywords by experts in the subject domain is integrated with formal computer language methodologies (ontologies).

In practice, however, it is difficult to characterize completely and precisely a technology using only a limited number of keywords. The larger and more inclusive is the choice of keywords, for example, including all synonyms of a given term, the greater the risk of finding unrelated patents due to polysemy or usage of the word in several other industries. Moreover, assignees may use (inadvertently or intentionally) different terminology to label the same technical concept. The list of variants is not known a priori.

Often inventions are described in ways that defy the precise qualification by means of keywords. Or the same functions are described differently, so leading to the publication of different keywords.

Finally, the labeling with keywords may miss important information. As a matter of fact, even the most obvious keywords may not be present in patent documents. For example, the CPC subgroup F04D 19/042 is about turbomolecular vacuum pumps, and yet there are 52 documents classified in F04D 19/042 that do not contain the term turbomolecular pump or any other variation of such an expression (source: http://www.wipo.int/meetings/en/details.jsp?meeting_id=39303).

40.3.4 Full Name of Assignees (Companies or Research Centers)

It is usually quite difficult to start with a complete list of companies and/or research centers that may be designated as assignees of patents. This is even more so in rapidly growing sectors, due to the massive entry of newcomers, as well as frequent mergers and acquisitions (M&A). A list of the full names of companies can be found in some industrial sectors in sources like industrial repositories, catalogues, trade associations, and associated websites. Complementary sources are the commercial database services.

However, even if a complete list were available, there are several limitations, some of which are similar to those commonly found in bibliometrics and scientometrics.

The well-known problem of harmonization of company names is pervasive: there are countless variations of company names to be found in patents. Harmonization efforts come into play, but they are still incomplete. In addition, companies try to hide their identity by assigning patents to subsidiaries whose corporate links are difficult to reconstruct, or even to their long term suppliers. In many cases, the assignees are inventors themselves, so the name of the company is not visible in the patent data. However, the inventors are employees or collaborators of a company. This information is not available in patent documents, so it must be inferred from other sources. As a matter of fact, the information may be difficult or impossible to reconstruct.

Finally, lists of companies are typically based on criteria for inclusion that refer mainly to the final products, i. e., are based on industry-sector criteria. This corresponds to the traditional notion that members of an industry are only those companies that actively compete in the product markets, or, more formally, those for which the cross-elasticity of product demand is non-zero. This notion was entirely appropriate in an innovation landscape in which there was a strong coherence between the technology owned or controlled by a company and its product portfolio. However, in a landscape of pervasive digital technologies and disruptive business models, this strict correspondence is not warranted. As an example, in emerging technologies, one often finds among assignees names of companies, usually large ones, coming from completely unrelated fields. This means that they are studying the technology. The extent to which they will develop products based on these patents, becoming new entrants and newcomer competitors, is not obvious at all.

40.3.5 Full Names of Inventors

The inventor record in the patent text is a source of crucial information. Many studies have been carried out by using lists of inventors, as well as their affiliation, country of origin, nationality, extracted from given sets of patents. An interesting example is the classification of inventors by country based on automatic tools of disambiguation, which assign a country or region with a certain probability given the frequency distribution of names and surnames.

Inventors are, however, physical persons. Contrary to names of companies and research centers, which create a universe in the order of magnitude of dozens of thousands, names of physical persons are in the order of millions. In addition, for companies there is an incentive to select corporate or brand names that are clearly distinguishable from competitors. This does not happen for physical persons. This means that issues of homonymy are cumbersome and may create lot of noise in data.

In addition, there is no validated list of inventors, for the time being.

For the convenience of the reader, the main limitations of existing patent search criteria and approaches are summarized in Table 40.2.

Table 40.2 Summary of the limitations of existing patent classifications

Search criteria	Limitations
IPC or CPC classes	• Large number of IPC codes and CPC entries generates cumbersome task of classification and information retrieval • Ambiguities of attribution • Classification errors • Strategic information hiding • Mismatch between speed of updating of classification and speed of emergence of new technologies • Cross-field technologies
Industry codes	• Classification errors • Research centers are not categorized by industry codes
Keywords	• Incomplete profiling of technologies using keywords • Polysemy • Ambiguity in the description of technologies by companies • Missing keywords
Full name of assignees (Name of companies or research centers)	• Cost of compiling lists of companies • Completeness of list of companies (new entry, M&A) • Name harmonization • Allocation of patents to subsidiaries with opaque corporate links and/or loyal and strategic suppliers in order to hide patent activity • Allocation of patents to individual inventors who are employees/collaborators in order to hide patent activity • Entry from unrelated industries
Full name of inventors	• Lack of completeness of inventor list • Homonymy

40.4 Functional Patent Classification: Three Case Studies

As mentioned in the previous chapters a functional patent classification (FPC) is based on the main functions performed by the technology, rather than on the inventive solutions or their potential applications. The functional approach allows overcoming most of the above-mentioned limits. One aspect that makes functions such a powerful tool is their generality and abstraction. Representing logical, physical, or teleological concepts, functions are neither domain specific nor domain dependent. As an example, separation, movement, and control are present in every technical domain, what changes is only the structure that realizes these general goals or effects. Therefore, functions can help the identification of connections or even the creation of bridges between distant technologies or industrial areas.

The connection may be found in the two time directions. Looking retrospectively, it is possible to start from a given present-day solution and explore inventions of the past belonging to different sectors, either to make more complete the positioning of a technology and the understanding of its evolution trajectory, or to widen the scope of infringement, opposition, or freedom to operate analyses.

Looking forward, on the contrary, the existing patent corpus can be used to provide inspiration for the inventions of the future, tackling a creative process called *crossover*, i. e., the adaptation of technical solutions from one field to another. The same approach can help in anticipating the evolution of transversal technologies.

Furthermore, the search of prior-art is very important in the every day practice of engineers, designers and IP professionals; however, the projection towards the future provided by crossover is probably even more important, since it leads to new technologies and businesses, and can provide valuable support to the strategic planning of companies and policy makers.

In the following sections, we will review some of the advantages of using FPC in a variety of directions. These include:

- Patent search
- Technology foresight
- Prior art
- Crossover analysis

In the first two case studies, we will show how adopting a functional reasoning and FPC allows a better retrieval of the patents:

- Related to a technological cluster, e. g., during a foresight activity.
- Related to a specific product that a company plan to patent and commercialize without infringing one or more existing IP (in the case of patent search and prior art analysis).

In the third case study, we will discuss how performing a patent search based on functional criteria permits us to identify different technologies that satisfy the same need and apply that to crossover activities, in order to support creative tasks in the conceptual design phase.

These applications are of interest for patent offices, patent attorneys, and patent analysts, as well as for entrepreneurs and venture capitalists, or researchers and analysts interested in technology and competitive intelligence.

40.4.1 Case Study No. 1: Patent Search

There are two main advantages of adopting a functional point of view when performing a patent search. The first is higher recall (in information retrieval; the term *recall* indicates a percentage parameter representing the completeness of a given target document set; in the present case, it gives the fraction of relevant patents that have been actually retrieved over the total amount of existing relevant inventions). Quite often, relevant patents are filed under IPC/CPC classes different from that of the starting patent application, and traditional queries are usually not able to retrieve them, either because they rely too much on the IPC/CPC patent classification, or because the keywords used are too domain dependent. Even similarity search, based on semantic technologies, usually fails in this task, since it still bases its internal representation on the specific terminology of the initial example.

The second advantage is that finding solutions coming from different fields is often unexpected and, therefore, offers additional weapons in the IP dialectics (for example, in patent litigation or opposition), similar to the possibility of utilizing non-patent literature. Moreover, the reverse is also true, that is, using the functional approach it is possible to detect patents that have been hidden in classes that are far away from the obvious one, either for defensive or offensive purposes.

In the foresight activity of the biomedical industry commissioned in 2017 by Toscana Life Science, a non-profit organization in the support of biomedical research and acceleration of startup companies, the starting point was the creation of the set of relevant patents.

The biomedical field has been clustered in 12 areas, defined at high level by using functional verbs that identify the main action performed by the technologies belonging to each area. For example, in the field of surgery, instead of listing individual technologies such as scalpel or cutting laser, we defined the cluster in terms of the main function, i. e., to separate/cut the tissues of a patient. Table 40.3 shows the main functions identified in the exercise. This segmentation is not intended to be exhaustive; it addresses the main ares of interest of the client. It gives a hint to the search strategy that the functional classification suggests.

Let us consider the cluster of technologies which function is to support the motor functions (listed as number 5 in Table 40.3). It contains the products and devices used for the rehabilitation and the aid of the mobility of a patient, such as crutches, wheelchairs, or training equipment.

Taking advantage of the functional dictionary to support the *functionalization* of the search, i. e., considering all possible variants of the functional concepts to be retrieved, we identified in this functional class a global patent set of 133 197 documents, belonging to 45 976 patent families, filed worldwide from 1900 to 2015.

Only a tiny fraction of these patents were filed in the region supporting the study (Tuscany). From the above set, in fact, 267 individual patents filed by assignees localized in the region have been found. They belong to 42 patent families, and their filing date is after 1985. Focusing on this small sample, it appears that some of the patents identified would not have been found, had we used the search criteria listed above.

For example, the application US2006113846_A1 (*Mechanism of motor reduction with variable rigidity*

Table 40.3 Classification of patents for medical devices adopting the similarity of functions performed as criterion for clustering

Segmentation of the technologies according to performed function	
1.	Technologies to remove/separate material
2.	Technologies to provide stimulus to the body
3.	Technologies to measure physiological parameters
4.	Technologies to detect images
5.	Technologies to support the motor functions
6.	Technologies to reach a specific part of the body
7.	Technologies to collect samples
8.	Technologies to pre-process samples
9.	Technologies to increase sample quantity
10.	Technologies to allow interaction of samples and reagents
11.	Technologies to detect signals from body fluid and tissue
12.	Technologies to sterilize the devices

and rapidly controllable) is classified under the IPC groups B25J9/02, F16H19/06 and H02N3/00. These groups are labeled, respectively, *Manipulators positioned in space by hand*, *Gearings comprising essentially only toothed gears or friction members and not capable of conveying indefinitely-continuing rotary motion* and *Generators in which thermal or kinetic energy is converted into electrical energy by ionisation of a fluid and removal of the charge therefrom*. If we had conducted the search using just those IPC classes for which the definition matches the concept of rehabilitation devices or mobility aids (that is, A61F,- *Medical or veterinary science; Hygiene; filters implantable into blood vessels; prostheses; devices providing patency to, or preventing collapsing of, tubular structures of the body; orthopaedic, nursing or contraceptive devices* or A61G,- *Medical or veterinary science; Hygiene; Transport, personal conveyances, or accommodation specially adapted for patients or disabled persons*), we would have not identified the above relevant US application, since it is classified under classes apparently unrelated to the biomedical field.

Rather often, the assignees and the inventor of a patent overlap. This is generally true for US applications, since in that jurisdiction there is the presumption that the inventor is the initial owner of a patent or patent application. Sometimes the inventor is, indeed, a single professional working on his/her own. More frequently, however, particularly in some industries, the inventor is an employee in the R&D department of a company and, by contract, the owner of the intellectual property is the company, not the inventor. In certain cases, the re-assignment to the legal entity from the physical person (inventor) to the company may still be in progress. In other cases, however, companies intentionally leave individual inventors as assignees, in order to hide the invention from competitors. Therefore, if these patents were searched using the names of the companies active in the industry, they would not be retrieved. For example, in the patent IT1252816_B (*Reinforced cotyle for hip joint prosthesis*) Mr. Massimo Giontella is both the assignee and the inventor. We started a search on other documents and discovered that this inventor works for a company (MP srl), and that this company owns several patents that refer to devices for the support of the motor functions. Had we searched for the standard criteria listed above, we would not have been able to reconstruct this hidden connection.

Another interesting remark about the above-mentioned company, MP srl, is that it performs mechanical manufacturing. Indeed, from its website it is not possible to infer that it produces biomedical equipment. For this reason, it would be difficult to retrieve its patents relying on the assignee information only, since it is not listed in any company list in the biomedical industry.

Similarly, we identified an assignee whose industrial classification was *Integrated engineering design services* (Ateco code 71.12.2 in the Italian industry classification). This industrial classification is too generic to infer any relatedness to the medical device industry. Yet it is the classification used for Prensilia, a university spinoff company, whose patent EP2653137_A1 (*Self-contained multifunctional hand prosthesis*) is clearly relevant to the biomedical industry. The patents of Prensilia would have been missed if the query had been based on industry classification only.

Finally, industrial classifications do not cover universities and research institutes and centers. In our case, as many as 34 patents related to the motor functions are assigned to Scuola Superiore Sant'Anna, a university institution. Thanks to the functional approach, it is, therefore, possible to find documents that do not have explicit reference to known assignees. In addition, it is also possible to elaborate on the relations between the technologies of interest and the strategic orientation of companies that do not appear in the core of the industry, and, therefore, are not under the regular scrutiny of competitors.

To sum up, in using the Functional Dictionary illustrated above, the levels of recall and precision were extremely high, by the standards adopted in the computational linguistics community (in information retrieval, the term *precision* indicates the fraction of relevant documents contained in the retrieved set; in the present case, such percentage parameter estimates how many of the patents in a given target patent-set do pertain to the technical area of interest). Functional thinking in general allows finding results that would have been missed otherwise, both by traditional patent search methods and by relying on pre-existing knowledge.

A final comment about the application of functional classification to the field of technology foresight is in order. Here, the functional approach is extended along the *time* dimension. The functional representation of technologies supports the identification of technical trends that project into the future the evolution of solutions, beyond the existing ones. This is a powerful counterbalance to the tendency of experts to reason of future technologies in terms of extensions of already existing solutions. Following the functional approach, the technology foresight may lead to the prediction of forthcoming solutions that fulfill the needs and goals emerging from the analysis, or, stated more formally, the functions of interest [40.93]. In other words, by investigating functions not properly addressed by existing solutions, as well as by extrapolating trends well known from the theory of functional analysis, it is possible to identify the directions along which the next innovative steps will take place. In addition, the functional approach allows the early identification of the potential failures of inventions. Failures can, in fact, be conceptualized as *negative* functions. A functional representation allows early detection of the areas in which the promised deliveries of benefits are likely to be frustrated [40.94].

40.4.2 Case Study No. 2: Prior Art and Out-of-Field Citations

The advantages of retrieving solutions coming from different fields were already pointed out in the previous section. There, the discussion was on search in general, but the same is true for the specific case of prior art search.

However, for anteriority search an objection can arise. How far apart (from the technical point of view) can two inventions be, so that one can still be considered a legitimate prior art of the other? Indeed, one may object that, in principle, there might be a threshold over which two solutions are so different that they can hardly be considered by a person skilled in the art to share a similar inventive step, even if they perform a similar function. However, there is no common agreed upon definition of an objective or measurable *distance* between artifacts that would allow setting such a threshold in a clear way. The judgement about the degree of similarity is usually left to the sensibility and experience of the IP professional. In addition, as a matter of fact, out-of-field citations are, indeed, used by patent examiners, patent attorneys, and companies' IP professionals.

To investigate the degree of usage/retrieval in prior art searches of solutions coming from external sectors, we used a set of over 200 000 patent applications, belonging to the biomedical sector and coming from several jurisdictions, which we had carefully selected for a previous study, and looked at the backward citations. For all data on citations and on search reports, we refer to the European Patent Office's PATSTAT service [40.95].

We assumed a very simple metric to compare any given application with its citations: two documents are considered pertaining to different sectors only if they have a different IPC/CPC class, i. e., if the first three characters of their IPC/CPC code are different. Any difference in the subsequent characters has no relevance for the present purposes. Such a metric, relying on the IPC classification tree, the criticalities of which we have already highlighted, is probably not the most accurate for an in-depth one-to-one comparison but can be easily automated to process large amount of documents, and the results are reliable on a statistical basis.

The study of the above-mentioned patent set led to some interesting results.

First, out-of-field citation is quite common; in almost one out of two patent applications (46%), the examiner cited in his/her search report at least one document belonging to a different sector. (Note that we restricted our analysis to citations made by patent office examiners and third parties only, neglecting the citations made by the applicant themselves. Please also note that the above percentage can be slightly overestimated because some documents have multiple IPC attributions, which may be both in-field and out-of-field.)

Out of almost 4 million backward citations from examiners for the whole set, 32% have a different IPC class with respect to the starting application (again, the exact percentage may be a bit lower when taking into account multiple attributions).

Second, even given the above, examiners very rarely rely on out-of-field citations only. In the various search reports of patent office examiners, around 10 000 applications were found to present prior art that would compromise the validity of one or more claims (X or Y categories of citation according to European Patent Office's convention: category X is applicable where a document is such that when taken alone, a claimed invention cannot be considered novel or cannot be considered to involve an inventive step; category Y is applicable where a document is such that a claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art). Of those, about half still presented at least one out-of-field citation, but only 0.3% (29 out of 10621) had *only* out-of-field citations. Even if we included non-invalidating citations (A category and similar), we reach only 0.9% of applications with out-of-field citations only.

Third, out-of-field citations are relatively more important in patent opposition (an opposition occurs when a third party challenges the validity of a patent; data for oppositions can also be found using the PATSTAT service). We found only 153 patent applications within the set that received an opposition. However, the percentage of documents opposed using out-of-field citations only now rises to 4.6% (7 out of 153), i. e., more than ten times the examiner's case.

We now turn to specific examples. Consider, for example, patent application EP1943975 (A1), about an *Holder for storage of surgical or medical equipment with filling template* and assigned to IPC subclass A61B (diagnosis, surgery, identification; a drawing of the invention can be seen in Fig. 40.1a). The only critical prior art (it received the X category) cited by the examiner during his/her search is US5379887 (A), about a *Method and apparatus for managing sewing machine spare parts* and assigned to subclasses B25H (workshop equipment) and D05B (sewing) (Fig. 40.1b). Although the application sectors are very different, the two documents obviously share the same main function, i. e., storage of objects. Reading the patents it is clear that they also share the additional function of displaying the correct position of objects within the box to the user. Similar functions often imply similar solutions, and indeed, as pointed out by the examiner, both patents recur for the display function to a template fixed to the lid.

As for EP 1479353 (A1) instead, a *Control panel* for electro-surgery devices, also filed under subclass A61B, the examiner has found only in-field prior art, such as for example patent DE3923024 (A1) about an *Electrosurgical apparatus with operating, display and safety device*. On the contrary, the patent application received an opposition citing the following three documents:

- DE10022588 (A1) an Electronic device under H04M (telephonic communication)
- de19951100 (A1) an Operating element, filed under H01H (electric switches) and with a clear automotive application
- WO0073867 (A1), an Indicator for a robotic machine, filed under various classes including A47L (domestic washing or cleaning) and concerning a robotic vacuum cleaner.

The documents belong to different sectors, yet they all perform the control and display functions in a similar way.

As a further example consider finally EP1670371 (A1), a *Transport device for sterile media* in A61B; the examiner cited, for example, the *Fluid jet blood sampling device and methods* of US 20020045912, still in A61B, but a competitor filed an opposition citing instead the *Flow control system for liquid chromatographs* of US4137011 (A) under, among others, F04B (positive-displacement machines for liquids; pumps).

Identifying out-of-field citations may be crucial for supporting patent litigation or for defending the competitive position against competitors. A strategy often adopted by attorneys that oppose a patent is to invoke the so called *general common knowledge*: if a solution is adopted in other industries, one should infer that it is largely known. It is, therefore, of crucial importance to carry out an extensive out-of-field search in order to anticipate potential arguments for opposition. Indeed, in a case we studied, there were similar solutions in at least seven (sic) different industries.

40.4.3 Case Study No. 3: Functional Crossover in Food Container Sterilization

As much as functions highlight connections between existing solutions in different sectors, they can be used to create a bridge to reach the inventions yet to be invented, thus fostering the innovation process. Patents can be a very interesting source of ideas to support

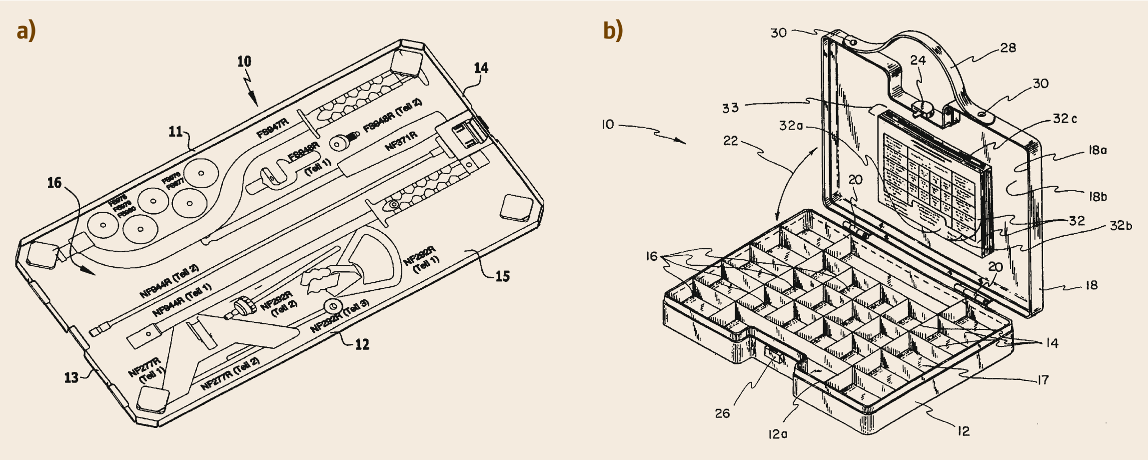


Fig. 40.1a,b Example of similarity of functions in patents from distant IPC classes: **(a)** comes from patent EP1943975 and represents a holder for surgical instruments (after [40.96]), **(b)** from patent US5379887, refers to an holder for sewing machine parts (after [40.97]). The two inventions belong to different industrial areas yet they perform the same function and indeed share also many features and part of the inventive step, as detailed in the text

the activities of inventors and designers in the concept design stage of the new product development (NPD) process. Indeed, understanding what has been created by others can spark creative solutions, in the form of variants or new combinations.

Going further, it is possible to use a technology traditionally developed in one industry to satisfy the needs of users in totally different fields of applications. This goes under the name of *crossover*, and it is a well-known way in which inventions are generated, consciously or not. For example, biomimetic is just a type of crossover, and while it is now a design discipline on its own, humans have always taken inspiration from nature for new inventions.

Crossover requires analogical reasoning, that is, the ability to identify the similarity between the deep structure of problems, beneath the surface of differences [40.98–100]. People capable of analogical reasoning discover similarity where ordinary people see only semantically irreducible problems.

Functional analysis offers a systematic approach to identify similarity across distant industries and products. It builds up abstract representations of the goals of products and technologies, that cut across existing solutions described in structural terms. In fact, the same functions may be found in completely different industries. Harnessing the functional approach coupled with a proper mining of the patent corpus is the most effective way to generate crossovers [40.85, 101, 102]. Several heuristics can be used for this purpose, such as the search for variants, the use of the same physical principles for different functions, the systematic search for synonyms and antonyms, and the like.

We applied functional analysis to the field of food container sterilization. The goal was to identify novel technologies, outside the focal industry. This challenge could not be addressed by relying on any of the search criteria discussed above; no patent classification, no list of companies or inventors, no industry classification, and no keywords were available, and if they had been available, they would not have allowed the discovery of the same result.

The preliminary stage was the formal definition of the main functions of a food container sterilizer (i. e., the destruction or removal of bacteria and other organisms harmful to humans). It is crucial that the functions come to be described in a clear and formal way. This requires a good understanding of the functional paradigm and can benefit from the use of a complete functional dictionary.

The full scale functional representation was then projected on the patent corpus in order to find those technologies that perform the functions, without imposing any restriction on the industrial sector. Following this approach, we identified as many as 50 patents about systems to sterilize materials and surfaces, outside the focal patent classes and industry classifications. In turn, these documents have been classified according to the physical effect underlying the patented technology, such as for example x-rays, gamma rays, plasma, ultrasounds, chemical agents, and so on (the latter classification can be performed in an automated way if a database of physical effects is available).

Table 40.4 shows a sample of results from this analysis.

Table 40.4 Sample of patents identified with an FPC in the field of food container sterilization

Cluster	UID	Title	Filing date	Assignee	Technical sectors
Magnetic field	US4524079_A	Deactivation of microorganisms by an oscillating magnetic field	1983	Maxwell Laboratories	Instruments/medical technology Chemistry/food chemistry Chemistry/biotechnology Chemistry/environmental technology
Micro waves + steam	WO9729016_A1	A method and an apparatus for surface sterilizing items and a system suitable for sterilizing bottles	1997	Clean-Pack Group	Mechanical engineering/handling
Alternate pressure	US6966345_B2	Method for durability treatment of a pumpable material as well as a device therefor	2003	Flow Holdings Sagl	Instruments/medical technology Mechanical engineering/handling Chemistry/food chemistry
Gamma rays	WO0043049_A1	Gamma-irradiation sterilized polyethylene packaging	1999	Pharmacia & Upjohn Company	Instruments/medical technology Mechanical engineering/handling
Ultrasounds	EP2550867_A1	Method for and device for control of microbes in food materials by means of vacuum and resonant ultrasound treatment	2011	University Of Miyazaki/Kaijo Corporation	Chemistry/food chemistry

In Table 40.4 we list a few patents that were identified through the functional approach. They are also classified according to the correspondence between IPC classes and technical sectors developed by *Schmoch* and co-authors [40.103, 104]. It is clear from the table that highly relevant patents are found in industries and technical sectors that have no proximity to the food or packaging industries, such as, for example, medical technology.

After the identification of these patents, it was possible to set up brainstorming sessions aimed at exploring the underlying inventive principles and their relevance for the sterilization of food containers. This activity led to the validation of a large number of product concepts: as many as 55. These concepts were then subject to a process of screening and refining, until a small number was selected for implementation.

40.5 Conclusions and Future Research

The notion of function is at the core of the patent system. However, the legal and economic doctrine of patents, as well as professional practice, have largely ignored the theoretical and empirical developments of this notion in fields such as engineering design and design theory.

It is time to make an effort to put this notion at the core of analysis and practice. We have shown that the theoretical treatment of the notion is now mature, from a philosophical and epistemological point of view, as well as in engineering disciplines. These conceptual developments offer a robust background for a systematic analysis of the notion of functions in the legal and economic doctrine of intellectual property. In turn, this might offer ground for more systematic and formal procedures of patent search carried out at patent offices.

We have also shown that the recent and impressive developments in computational linguistics and the automatic treatment of texts open the way for new applications.

In this chapter, we have suggested the integration between current approaches to patent classification and the functional classification approach. It is clear that a full scale, *pure* functional classification of all existing patents is a long term goal, requiring further research over many years. However, a promising intermediate step might be to compare existing classifications with functional classification in limited, controllable, new areas of technology that require dedicated efforts of updating. Are the current approaches to classify patents, say, in the field of Industry 4.0, appropriate? Or in the field of FinTech? It would be useful to develop a formal framework for the comparison of alternative approaches, based on well-defined metrics drawn from computational linguistics and from graph theory (e. g., precision, recall, predictive power, number, and share of relevant out-of-the field citations identified, and measures of distance in the classification graph).

Another long term goal, which is, however, made realistic by the current developments in computational linguistics, is the definition of formal measures of technological distance and its semi-automatic computation.

Keeping the full scale functional classification as a long term goal, other short term applications are already very promising. In the field of patent search and patent analysis, the functional approach allows us to overcome the limitations of existing classifications, by identifying several relevant inventions that would remain hidden otherwise. Applications to patent search will prove valuable in prior art analysis, freedom to operate, and litigation. Patent offices might find it useful to incorporate it in their routine procedures.

The functional approach offers new perspectives in fields of analysis that use patent datasets for a variety of purposes. It is a powerful tool for the profiling of emerging technologies, beyond existing technology or industry boundaries. It allows the identification of lateral opportunities, analogical solutions, and crossover applications in innovation management. It supports a systematic projection of technologies in the future, in studies of technology foresight, mitigating the cognitive and motivational biases of experts.

A promising direction is the use of large scale functional dictionaries, based on deep engineering domain knowledge, coupled with powerful linguistic tools. Given the success in developing large scale dictionaries based on functions, the same approach should be followed in the effort to reach saturation in dictionaries that deal with stakeholders/users, advantages and disadvantages, and physical descriptions of structures and behaviors.

A large research agenda is therefore open.

Acknowledgments. The authors would like to thank Dario Brugnoli, Giovanni De Santis, Tommaso Pavanello, and Corrado Taviani for support in the development of the case studies.

References

40.1 V. Hubka, W.E. Eder: *Theory of Technical Systems* (Springer, Berlin 1988)

40.2 A. Bonaccorsi, J. Calvert, P.B. Joly: From protecting texts to protecting objects in biotechnology and software. A tale of changes of ontological assumptions in intellectual property protection, Econ. Soc. **40**(4), 611–639 (2011)

40.3 A. Ariew: Platonic and Aristotelian roots of teleological arguments. In: *Functions. New Essays in the Philosophy of Psychology and Biology*, ed. by A. Ariew, R. Cummins, M. Perlman (Oxford University Press, Oxford 2002) pp. 7–32

40.4 W.F. Hood: The Aristotelian versus the Heideggerian approach to the problem of technology. In: *Philosophy and Technology. Readings in the Philosophical Problems of Technology*, ed. by C. Mitcham, R. Mackey (Free Press, New York 1983) pp. 347–363

40.5 F.J. Ayala: Teleological explanations in evolutionary biology, Philos. Sci. **37**, 1–15 (1970)

40.6 E. Mayr: *The Growth of Biological Thought. Diversity, Evolution and Inheritance* (Belknap Press, Cambridge 1982)

40.7 R. Amundson, G.V. Lauder: 'Function without purpose': The use of causal role function in evolutionary biology, Biol. Philos. **9**, 443–469 (1994)

40.8 M. Ruse: Evolutionary biology and teleological thinking. In: *Functions. New Essays in the Philosophy of Psychology and Biology* (Oxford University Press, Oxford 2002) pp. 33–59

40.9 E. Sober (Ed.): *Conceptual Issues in Evolutionary Biology* (MIT Press, Cambridge 1984)

40.10 N. Eldredge: *Reinventing Darwin. The Great Debate at the High Table of Evolutionary Theory* (Wiley, New York 1995)

40.11 M. Pigliucci, J. Kaplan: *Making Sense of Evolution. The Conceptual Foundations of Evolutionary Biology* (University of Chicago Press, Chicago 2006)

40.12 E. Nagel: *The Structure of Science* (Harcourt Brace, New York 1961)

40.13 C.G. Hempel: *Aspects of Scientific Explanation* (Free Press, New York 1965)

40.14 S.J. Gould, R.C. Lewontin: The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme, Proc. R. Soc. B **205**(1161), 581–598 (1979)

40.15 S.J. Gould, E.S. Vrba: Exaptation. A missing term in the science of form, Paleobiology **8**(1), 4–15 (1982)

40.16 E. Vrba, S.J. Gould: The hierarchical expansion of sorting and selection: Sorting and selection cannot be equated, Paleobiology **12**, 217–228 (1986)

40.17 S.J. Gould: *The Structure of Evolutionary Theory* (Belknap Press, Cambridge 2002)

40.18 L. Wright: Functions, Philos. Rev. **82**, 139–168 (1973)

40.19 R. Cummins: Functional analysis, J. Philos. **72**, 741–765 (1975)

40.20 R.G. Millikan: *Language, Thought, and Other Biological Categories. New Foundations for Realism* (MIT Press, Cambridge 1984)

40.21 P. McLaughlin: *What Functions Explain. Functional Explanation and Self-Reproducing Systems* (Cambridge University Press, Cambridge 2001)

40.22 A. Ariew, R. Cummins, M. Perlman (Eds.): *Functions. New Essays in the Philosophy of Psychology and Biology* (Oxford University Press, Oxford 2002)

40.23 T. Lewens: *Organisms and Artifacts. Design in Nature and Elsewhere* (MIT Press, Cambridge 2004)

40.24 U. Krohs, P. Kroes (Eds.): *Functions in Biological and Artificial Worlds. Comparative Philosophical Perspectives* (MIT Press, Cambridge 2009)

40.25 P.E. Vermaas: The physical connection. Engineering function ascription to technical artefacts and their components, Stud. Hist. Philos. Sci. **37**, 62–75 (2006)

40.26 P.E. Vermaas, W. Houkes: Technical functions. A drawbridge between the intentional and structural natures of technical artefacts, Stud. Hist. Philos. Sci. **37**, 62–75 (2006)

40.27 A. Bonaccorsi: A functional theory of technology and technological change. In: *Handbook on the Economic Complexity of Technological Change*, ed. by C. Antonelli (Edward Elgar, Cheltenham 2011)

40.28 G. Pahl, W. Beitz: *Engineering Design: Systematic Approach*, 2nd edn. (Springer, Berlin 1996)

40.29 K. Wallace, L. Blessing: Introduction. In: *Engineering Design. A Systematic Approach*, 3rd edn., ed. by G. Pahl, W. Beitz, J. Feldhusen, K.H. Grote (Springer, Berlin 2007)

40.30 V. Hubka, M. Andreasen, W. Eder, P. Hils: *Practical Studies in Systematic Design* (Butterworth, London 1988)

40.31 G.S. Altschuller: *Creativity as an Exact Science: Theory of the Solution of Inventive Problems* (Gordon and Breach, Amsterdam 1987)

40.32 H.A. Simon: The architecture of complexity, Proc. Am. Philos. Soc. **106**(6), 467–482 (1962)

40.33 H.A. Simon: *The Sciences of the Artificial* (MIT Press, Cambridge 1969)

40.34 A. Newell, H.A. Simon: *Human Problem Solving* (Prentice-Hall, Englewood Cliffs 1972)

40.35 A. Newell: *Unified Theories of Cognition* (Harvard University Press, Cambridge 1990)

40.36 E. Subrahmanian: *An Intelligent, Computer-Based Tutor for Bottlenecks in Engineering and Science. Final Report* (Carnegie Mellon University, Pittsburgh 1987)

40.37 S. Finger, J.R. Dixon: A review of research in Mechanical engineering design. Part I: Descriptive, prescriptive and computer-based models of design processes, Res. Eng. Des. **1**, 51–67 (1989)

40.38 R.F. Coyne: *Planning in Design Synthesis: Abstraction-Based LOOS (ABLOOS), Ph.D. Thesis* (Carnegie Mellon University, Pittsburgh 1989)

40.39 R.F. Coyne, U. Fleming, P. Piela, R. Woodbury: Behavior modeling in design system development. In: *CAAD Futures '93, Pittsburgh* (1993) pp. 335–354
40.40 R.F. Coyne, E. Subrahmanian: Computer supported creative design: A pragmatic approach. In: *Modeling Creativity and Knowledge-Based Design*, ed. by J.S. Gero, M.L. Maher (Lawrence Erlbmaum, Hillsdale 1993) pp. 295–327
40.41 R.H. Sturges: A computational model for conceptual design based on function logic. In: *Artificial Intelligence in Design '92*, ed. by J.S. Gero, F. Sudweeks (Springer, Dordrecht 1992) pp. 757–772
40.42 C. Tong, D. Sriram: *Artificial Intelligence in Engineering Design* (Academic Press, Boston 1992)
40.43 D. Sriram: *Intelligent Systems for Engineering* (Springer, Berlin 1997)
40.44 D.C. Brown, B. Chandrasekaran: *Design Problem Solving. Knowledge Structures and Control Strategies* (Pitman, London 1989)
40.45 V. Goel, P. Pirolli: The structure of design problem spaces, Cogn. Sci. **16**, 395–429 (1992)
40.46 A. Little, K. Wood, D. McAdams: Functional analysis: A fundamental empirical study for reverse engineering, benchmarking and redesign. In: *Proc. ASME Des. Theory Methodol. Conf., Sacramento* (1997)
40.47 R. Stone, K. Wood: Development of a functional basis for design, J. Mech. Des. **122**(4), 359–370 (2000)
40.48 J. Hirtz, R. Stone, D. McAdams, S. Szykman, K. Wood: A functional basis for engineering design: Reconciling and evolving previous efforts, Res. Eng. Des. **13**, 65–82 (2002)
40.49 R.B. Stone, A. Chakrabarti (Eds.): Engineering applications of representations of function, J. Artif. Intell. Eng. Des. Anal. Manuf. **19**(2/3) (2005)
40.50 M.M. Van Wie, C.R. Bryant, M.R. Bohm, D. McAdams, R.B. Stone: A model of function-based representations, Artif. Intell. Eng. Des. Anal. Manuf. **19**(2), 89–111 (2005)
40.51 P.E. Vermaas: The functional modeling account of Stone and Wood: Some critical remarks. In: *Int. Conf. Eng. Des., ICED* (2007)
40.52 Y. Umeda, H. Takeda, T. Tomiyama, H. Yoshikawa: Function, behaviour, and structure, Appl. Artif. Intell. Eng. V **1**, 177–194 (1990)
40.53 J.S. Gero: Design prototypes: A Knowledge Representation schema for design, AI Magazine **11**(4), 26–36 (1990)
40.54 R.D. Coyne, M.A. Rosenman, A.D. Radford, M. Balachandran, J.S. Gero: *Knowledge-Based Design Systems* (Addison-Wesley, Reading 1990)
40.55 M.A. Rosenman, J.S. Gero: Purpose and function in design: From the socio-cultural to the technophysical, Des. Stud. **19**(2), 161–186 (1998)
40.56 S. Gero, U. Kannengiesser: The situated function–behaviour–structure framework. In: *Artificial Intelligence in Design '02*, ed. by J.S. Gero (Kluwer, Dordrecht 2002) pp. 89–104
40.57 D. Brown: Functional, behavioral and structural features. In: *ASME Des. Eng. Tech. Conf. Proc* (2003), DET2003/DTM-48684
40.58 D.C. Brown, L. Blessing: The relationship between function and affordance. In: *ASME 2005 Des. Theory Methodol. Conf., Long Beach* (2005), DETC2005-85017
40.59 G. Cascini, L. Del Frate, G. Fantoni, F. Montagna: Beyond the design perspective of Gero's FBS Framework. In: *Design Computing and Cognition '10*, ed. by J.S. Gero (Springer, Dordrecht 2010) pp. 77–96
40.60 G. Cascini, G. Fantoni, F. Montagna: Situating needs and requirements in the FBS framework, J. Des. Stud. **34**(5), 636–662 (2013)
40.61 K. Dorst, P.E. Vermaas: John Gero's function-behaviour-structure model of designing: A critical analysis, Res. Eng. Des. **16**, 17–26 (2005)
40.62 P.E. Vermaas, K. Dorst: On the conceptual framework of John Gero's FBS-model and the prescriptive aims of design methodology, Des. Stud. **28**, 133–157 (2007)
40.63 Y. Umeda, M. Ishii, M. Yoshioka, Y. Shimomura, T. Tomiyama: Supporting conceptual design based on the function-behavior-state modeler, Artif. Intell. Eng. Des. Anal. Manuf. **10**, 275–288 (1996)
40.64 N.P. Suh: *Axiomatic Design: Advances and Applications* (Oxford University Press, New York 2001)
40.65 Y. Kitamura, T. Sano, K. Namba, R. Mizoguchi: A functional concept ontology and its application to automatic identification of functional structures, Adv. Eng. Inform. **16**(2), 145–163 (2002)
40.66 Y. Kitamura, M. Kashiwase, M. Fuse, R. Mizoguchi: Deployment of an ontological framework of functional design knowledge, Adv. Eng. Inform. **18**, 115–127 (2004)
40.67 Y. Kitamura, R. Mizoguchi: Ontology-based description of functional design knowledge and its use in a functional way server, Expert Syst. Appl. **24**, 153–166 (2003)
40.68 Y. Kitamura, R. Mizoguchi: Ontology-based systematization of functional knowledge, J. Eng. Des. **15**, 327–351 (2004)
40.69 M. Sasajima, Y. Kitamura, M. Ikeda, R. Mizoguchi: FBRL: A function and behavior representation language, Expert Syst. Appl. **10**(3/4), 447–471 (1995)
40.70 M.S. Erden, H. Komoto, T.J. van Beek, V. D'Amelio, V. Echavarria, T. Tomiyama: A review of function modeling: Approaches and applications, Artif. Intell. Eng. Des. Anal. Manuf. **22**, 147–169 (2008)
40.71 T. Montecchi, D. Russo, Y. Liu: Searching in cooperative patent classification: Comparison between keyword and concept-based search, Adv. Eng. Inform. **27**(3), 335–345 (2013)
40.72 D. Russo, T. Montecchi: A function-behaviour oriented search for patent digging. In: *Int. Design Eng. Tech. Conf. Comput. Inf. Eng. Conf. ASME* (2011)
40.73 T. Montecchi, D. Russo: FBOS: function/behaviour-oriented search, Procedia Eng. **131**, 140–149 (2015)
40.74 J. Yoon, K. Kim: Detecting signals of new technological opportunities using semantic patent

analysis and outlier detection, Scientometrics **90**, 445–461 (2012)
40.75 S. Choi, H. Park, D. Kang, J.Y. Lee, K. Kim: An SAO-based text mining approach to building a technology tree for technology planning, Expert Syst. Appl. **39**, 11443–11455 (2012)
40.76 H. Park, J.J. Ree, K. Kim: Identification of promising patents for technology transfers using TRIZ evolution trends, Expert Syst. Appl. **40**, 736–743 (2013)
40.77 G. Cascini: *System and Method for Automatically Performing Functional Analyses of Technical Texts*, European Patent EP1351156 (2002)
40.78 G. Cascini, F. Neri: Natural language processing for patents analysis and classification. In: *Proc. TRIZ Future 4th World Conf., Florence* (Firenze Univ. Press, Florence 2004)
40.79 G. Cascini, D. Russo: Computer-aided analysis of patents and search for TRIZ contradictions, Int. J. Prod. Dev. **4**(1/2), 52–67 (2007)
40.80 J.R. Curran, M. Moens: Improvements in automatic thesaurus extraction. In: *Proc. ACL-02 Workshop Unsuperv. Lex. Acquis., ULA '02, Philadelphia*, Vol. 9 (2002) pp. 59–66
40.81 C.V. Trappey, H.-Y. Wu, F. Taghaboni-Dutta, A.J.C. Trappey: Using patent data for technology forecasting: China RFID patent analysis, Adv. Eng. Inform. **25**(1), 53–64 (2011)
40.82 A. Bonaccorsi, G. Fantoni: An expanded functional base. In: *Proc. Int. Conf. Eng. Design ICED07, Paris* (2007)
40.83 R. Apreda, G. Fantoni, A. Bonaccorsi: Functional vector space. In: *Proc. Int. Conf. Eng. Des. ICED09, Palo Alto* (2009)
40.84 G. Fantoni, R. Apreda, D. Gabelloni, A. Bonaccorsi: Do functions exist? In: *Proc. Int. Conf. Eng. Des., ICED11 Copenhagen* (2011)
40.85 D. Gabelloni, R. Apreda, G. Fantoni: On the link between features and functions. In: *Proc. Int. Conf. Eng. Des., ICED11 Copenhagen* (2011)
40.86 D. Gabelloni, G. Fantoni: Designers' promises or users' expectations? In: *Proc. Int. Conf. Eng. Des., ICED13, Seul* (2013)
40.87 G. Fantoni, R. Apreda, F. Dell'Orletta, M. Monge: Automatic extraction of function-behaviour-state information from patents, Adv. Eng. Inform. **27**(3), 317–334 (2013)
40.88 F. Chiarello, G. Fantoni, A. Cimino, F. Dell'Orletta: Automatic user extraction from patents, World Patent Information, **54**, 28–38 (2018)
40.89 F. Chiarello, L. Trivelli, A. Bonaccorsi, G. Fantoni: Development of an enriched dictionary for semantic analysis of documents in the Industry 4.0 domain, Comput. Ind. **100**, 244–257 (2017)
40.90 World Intellectual Property Organization: International Patent Classification (IPC), http://www.wipo.int/classifications/ipc/en/
40.91 European Patent Office: Espacenet Patent Search, https://worldwide.espacenet.com/classification
40.92 F.P. Appio, R. Apreda, G. Fantoni, A. Martini: At the roots of an interdisciplinary radical innovation: preliminary insights from the patenting activity in the bioinformatics industry. In: *Proc. 15th Int. CINet Conf.*, 85–107 (2014)
40.93 R. Apreda, A. Bonaccorsi, F. dell'Orletta, G. Fantoni: Technology foresight based on functional analysis, Eur. J Futures Res. **1**, 4–13 (2016)
40.94 R. Apreda, A. Bonaccorsi, G. Fantoni, D. Gabelloni: Functions and failures: How to manage technological promises for societal challenges, Technol. Anal. Strateg. Manag. **26**(4), 369–384 (2014)
40.95 European Patent Office: PATSTAT https://www.epo.org/searching-for-patents/business/patstat.html
40.96 F.W. Oertmann, S. Schuster, Aesculap AG: Holder for storage of surgical or medical equipment with filling template EP patent 1943975 (2010)
40.97 R.F. Conley, MIM Industries Inc.: Method and apparatus for managing sewing machine spare parts, US Patent 5379887 (1995)
40.98 D. McAdams, K. Wood: Quantitative measures for design by analogy. In: *Proc. ASME Des. Theory Methodol. Conf* (2000), DETC2000/DTM-14562
40.99 J.S. Linsey, A.B. Markman, K.L. Wood: Design by analogy: A study of the *WordTree* method for problem re-representation, J. Mech. Des. **134**(4), 041009 (2012)
40.100 A.K. Goel, S.R. Bhatta: Use of design patterns in analogy-based design, Adv. Eng. Inform. **18**, 85–94 (2004)
40.101 G. Fantoni, C. Taviani, R. Santoro: Design by functional synonyms and antonyms: A structured creative technique based on functional analysis, Proc. Inst. Mech. Eng. Part B **221**(4), 673–683 (2006)
40.102 G. Fantoni, D. Gabelloni, J. Tilli: Concept design of new grippers using abstraction and analogy, Proc. Inst. Mech. Eng. Part B **227**(10), 1521–1532 (2013)
40.103 U. Schmoch, F. Laville, P. Patel, R. Frietsch: Linking technology areas to industrial sectors, Final report to the European Commission, DG Research (2003)
40.104 U. Schmoch U: Concept of a technology classification for country comparisons, Final Report to the World Intellectual Property Organisation (WIPO), revised November 2010, WIPO, Geneva (2010). http://www.wipo.int/export/sites/www/ipstats/en/statistics/patents/pdf/wipo_ipc_technology.pdf

Andrea Bonaccorsi

DESTEC
University of Pisa
Pisa, Italy
andrea.bonaccorsi@unipi.it

Andrea Bonaccorsi is Full Professor of Economics and Management at the School of Engineering of the University of Pisa. He has authored more than 60 articles in the most important journals in economics of science, technology and innovation. His main research interests are: economics of science (analysis of higher education institutions based on microdata, dynamics of science, research evaluation), and economics of technology (functional analysis, technology foresight, patent analysis).

Gualtiero Fantoni

Department of Civil and Industrial Engineering
University of Pisa
Pisa, Italy
g.fantoni@ing.unipi.it

Gualtiero Fantoni is Associate Professor of Advanced Manufacturing Systems and New Product Development at the Engineering Faculty of the University of Pisa. He received a PhD in Automation, Robotic and Bioengineering from the University of Pisa (2005). His research includes smart grippers in the field of robotics applied to manufacturing and assembly, technical text analysis, and smart programming tools for optimizing production and predictive maintenance.

Riccardo Apreda

Erre Quadro s.r.l.
Pisa, Italy
apreda.riccardo@gmail.com

Riccardo Apreda is Co-founder, former MD and currently Head of R&D of Erre Quadro Srl, a company specialized in software and services for IP valorisation. He received a PhD in Physics from the University of Pisa (2003). His research focuses on development of algorithms for analysis of technical texts; he is also interested in the theory of Engineering Design, in technology foresight and in methodologies to support the innovation process.

Donata Gabelloni

Erre Quadro s.r.l.
Pisa, Italy
d.gabelloni@gmail.com

Donata Gabelloni works at Erre Quadro Srl, Pisa, Italy, where she manages R&D projects on semi-automatic tools for patent analysis and on how to use data extracted from patents to map technologies and trends. She holds a PhD in Management Engineering from the University of Roma Tor Vergata. Her research interests are methodologies for new product development, functional analysis and semi-automatic analysis of technical documents, such as patents.

Part F

Patent Sy

Part F Patent System, Patents and Economics

41. Computer-Implemented Inventions in Europe

Peter Neuhäusler, Rainer Frietsch

The dispute between proponents and opponents of the patent system has been especially visible with regard to the patenting of computer programs. Different developments have resulted in the fact that there are large differences in the patent practices between the European Patent Office (EPO) and the U.S. Patent and Trademark Office (USPTO). While software *as such* is patentable at the USPTO, the EPO prohibits patenting of pure computer programs and only allows patenting of computer implemented inventions (CII).

In this chapter, we investigate the differences between the European and American patent systems with regard to patenting computer programs by also addressing the historical developments that have resulted in the national differences. Based on these considerations, a definition of CII is derived, which enables us to carry out empirical analyses.

By applying a conservative estimate, our results show that the share of CII filings at the EPO lies at around 25% at present, while at the USPTO a current margin of approximately 33% is reached. Thus, at least every fourth patent at the EPO and every third patent at the USPTO is a CII filing. In order to take account of the factual (technological and economical) relevance of computer-implemented inventions, we argue for clear rules with regard to patenting CII, as they are essential to reduce uncertainties and provide the relevant incentives for innovation.

41.1 Starting Points

Patents and other intellectual property rights are the pillars of every innovation system and provide substantial support for technology development and economic growth of national economies [41.1]. When issuing a patent, the state grants the patent holder a temporary monopoly—usually up to 20 years—on the rights to utilize and commercialize a technological solution. In return, the patent applicant needs to publish all the information about the underlying invention [41.2–5]. This is intended to support investments in new technologies and innovations on the one hand, and increase planning security for companies and research organizations on the other. The disclosure obligation hereby ensures that the knowledge generated in the innovation process is made available to the public in order to spur technological spill-over effects. The basic intention of the patent

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_41

system is thus to promote inventions and innovations within a national economy and to contribute to developing international competitiveness.

Proponents of the patent system emphasize its planning security, the clarity of rules, and the resulting incentives for innovation. Opponents of the system (or parts of it), on the other hand, state that the creation of temporary monopolies slows down innovative activities and prevents competition of the best technological solutions. It is argued that innovation capacities could be enhanced by repealing patents or at least changing parts of the system [41.6–11].

This dispute between advocates and critics of the patent system has been especially visible with regard to patents for computer programs (for a more detailed discussion on software patents see *Blind* et al. [41.12–14]). Similar disputes, albeit with slightly different arguments, can only be found in the field of genetic engineering. Some critics suggest that software inventions are not inventions in the basic sense but rather discoveries that are fundamentally excluded from patenting. Others argue that computer programs do not have a technological content (or a technological orientation) and thus want to exclude software from patenting. This has led to the fact that the patenting practices for computer programs differ across patent offices worldwide. At the USPTO, for example, software *as such* is patentable, while at the EPO, only *CII* can be patented, i. e., computer programs need to have a *technical nature* or generate a *technical effect* in order to be patentable. Consequently, at the dimension of technological content a gray zone between technology and software emerges, which is further enhanced by differing rules at different patent offices for defining software patents or CII and dealing with their patentability. This has become particularly apparent between European and American patent laws.

The objective of this chapter is to first of all to describe the pros and cons of patent systems from an economic perspective and discuss the differences between the European and American systems with regard to patenting software and CII. We will focus on patenting CII and software by also addressing the historical developments that have resulted in the national differences with regard to patentability. In a second step, we will provide a definition of CII and describe its operationalization for use in patent statistics. Finally, empirical trends in CII patenting are presented, and we will demonstrate the differences in the patenting practice for CII at the EPO and the USPTO. In sum, we try to obtain an overview of the current facts and backgrounds with regard to CII patenting to be able to understand their economic implications.

41.2 A Brief Introduction to the Economics of Intellectual Property Rights

One of the oldest questions of public innovation and technology policy is how to protect the results of innovation against uncontrolled use by third parties [41.15]. This is justified by the fact that innovation can be regarded as one of the key factors for economic growth on a micro as well as on a macro level [41.16–20]. The successful completion of an innovation process, however, is not a sufficient condition for attaining the expected benefits of innovation. Companies must also be able to appropriate their innovative results, i. e., they need to prevent competitors from imitating [41.21].

The reason for this lies in the nature of innovations or (technological) knowledge in general. Unlike traditionally produced and traded goods, (technological) knowledge represents public goods, i. e., it is non-rivaling and non-exclusive [41.5, 22]. The aspect of non-rivalry implies that increasing the number of users of technical knowledge does not limit its value, since anyone with the specific knowledge has the potential to achieve the same performance. The aspect of non-exclusiveness implies that third parties cannot (easily) be excluded from using certain goods. Therefore, third parties can reproduce knowledge without further marginal costs, which results in a suboptimal supply of (technological) knowledge, as the original expenditure for research and development (R&D) cannot be recouped [41.23].

Consequently, the production of knowledge and its commodification into an innovation suffers from market failure [41.24]. In the context of economic rationality, companies lack the incentives to invest in R&D. Thus, state intervention and the creation of suitable institutional framework conditions are necessary to enable the generation of private innovation rents and provide incentives for future (private) innovative efforts. The most important institutional structure to prevent this market failure and support the public goods of knowledge generation is the intellectual property rights (IPR) system. It guarantees the IP holder exclusive rights on the generated knowledge and a mechanism to pursue violations of their intellectual property, at least for a certain amount of time [41.15]). This can be understood as an incentive for innovators to develop new knowledge and new technologies. In exchange for this legal security, intellectual property rights are coupled with a disclosure obligation. All the information protected by a property right must be made publicly available after a certain time period (usually 18 months), allowing the diffusion of knowledge and potentially generating spill-over effects. Intellectual property rights are, thus, a state-guaranteed instrument that grants the innovator a monopoly for a specific time period in exchange for making the knowledge available to the general public [41.25].

41.2.1 Patent Law and Supranational Patent Systems

Basic patent law was already being applied in the eighteenth century in many countries and still counts as one of the major prerequisites for private sector innovativeness [41.15]. The protection of a patent is always limited to a national territory. If a company registers a patent in Germany, for example, the underlying invention is only protected in Germany. To obtain the same protection in the USA, a second patent with the same content has to be registered at the USPTO [41.5].

As the European market converged, however, a European patent procedure was introduced in 1978. Since then, it has been possible to apply for a patent at the EPO and designate countries for the patent to be forwarded after is has been granted. If the patent is granted, the patent protection is subsequently converted into national patents [41.5]. The European patent system can, thus, be regarded as a *system of systems* [41.26]. Besides patent law at the European level, however, each state that signed the European Patent Convention (EPC) still has its own national case law with regard to patents [41.26].

In addition to the EPO, there is the *Patent Cooperation Treaty* (PCT), which enables applicants to register their patents in a standardized procedure in all the countries that signed the PCT. The responsible authority is the World Intellectual Property Organization (WIPO) in Geneva. Unlike the EPO and other patent offices, the PCT procedure is only a filing procedure. After being filed at WIPO, patents are forwarded to national or supranational offices to be examined and (potentially) granted.

Patents are still mainly used as protection against imitation and for market security. However, the patent system's existence leads to possibilities to use patent protection for other—strategically motivated—purposes [41.27]. These strategic motives have been gaining importance since the beginning of the 1990s [41.28–31]. Following *Arundel* and *Patel* [41.32], all motives that go beyond protecting one's invention in order to appropriate the benefits based on this invention can be defined as strategic. The most frequent strategic motivation is blocking competitors by attempting to prevent other market players from using technical inventions in the same or neighboring fields of application. This is achieved, for example, by constructing what has become known as patent thickets. The consequence of strategic patenting is that the decision to patent has at least partially been decoupled from the technological necessities of protecting one's own invention against imitation by other players in the market.

41.2.2 Pros and Cons of the Patent System from an Economic Perspective

As mentioned in Sect. 41.2, intellectual property rights are an incentive mechanism of the state to promote the generation and diffusion of knowledge. However, so far, it has not been possible to empirically establish a clear causal link between stronger IPR and a growth of innovation [41.9]. Consequently, intellectual property rights have always been reviewed critically. The key question here is: Do intellectual property rights promote or obstruct innovation? Criticism is frequently formulated in the form of exceptions to the rule, i. e., limited to a type of innovation in specific technologies or based on the differences between small and large enterprises [41.9]. However, some of the critical considerations cannot simply be disregarded. The following section briefly outlines the positive and negative views on IPR and the economic arguments put forward by *Hahn* [41.9], *Guellec* [41.33], and *Mersch* [41.34], before we address the specific conditions for computer-implemented inventions in more detail.

The positive viewpoint usually starts with the already mentioned market failure when knowledge is being generated. There is no incentive for innovators to invest in knowledge production due to the non-excludability of the generated knowledge. This incentive has to be set in order to counteract insufficient investment in innovation in an economy [41.24]. It is further suggested that IPR (especially patents) also support the commercialization of inventions [41.35], by a) preventing imitation and acting as a signal to keep potential rivals out of the patent holder's own research fields or b) enable research to be conducted together with rivals (contractual agreements, licensing, etc.) [41.36].

Finally, due to the disclosure obligation, the generated knowledge is published relatively early after an invention has been made, which fosters sequential innovations and may lead to spill-over effects [41.37–41]. Furthermore, an invention is freely available to everyone after its right to protection has expired. In summary, it can be stated that intellectual property rights support knowledge generation, its commercialization, and its diffusion.

Critics of the patent system, on the other hand, usually present three partially interdependent arguments to demonstrate that intellectual property rights can also be obstructive to innovation. The first argument is aimed at the monopoly status of an IPR holder. The guaranteed monopoly—which is generally understood as an incentive to innovate—leads to an innovative output that is below a (socially) optimal level, since the monopoly holder has the possibility to set a price for an invention that lies (far) above its (estimated) costs. In this case, IPR would limit the diffusion of technologies and, therefore, also the diffusion of the generated knowledge ([41.9], based on [41.24]). A second line of argumentation that was decisively influenced by *Bessen* and *Meurer* for the US patent system [41.6–8] is that the costs of the patent system exceed its benefits. Costs mainly arise due to the legal costs of patent disputes that are caused, among other things, by patent thickets, i. e., unclear patent boundaries, which lead to the occurrence of overlapping patent rights. This argument is particularly valid when the behavior of private actors leads to significant social (legal) costs that have to be borne by the general public. The third argument concerns the already mentioned patent thickets and is of particular interest for new technologies, especially if these are complex ones that rely heavily on earlier technological developments, such as is the case, for example, in the field of software or biotechnology [41.9, 42]. Because this argument is frequently made in connection with computer-implemented inventions, the next chapter addresses this criticism in more detail.

41.2.3 Patent Thickets and the Tragedy of the Anti-Commons

Especially in high technology industries such as communications, computers, semiconductors, or biotechnology, where products are frequently protected by a large number of patents, and the innovation process is strongly sequential and cumulative, patents can lead to coordination problems that can result in significant costs [41.11]. A coordination problem occurs, for example, if a very broad *pioneer patent* is granted in an early phase of technology development. Although coordinated sequential research can still be carried out [41.35], this kind of patent can also represent a barrier to further innovation [41.9], since, at the very least, expenses for coordination and information exchange have to be carried. Although these costs can partly be minimized by early settlements, contracts, and license agreements, a pioneer patent can also have the effect that subsequent research results violate this patent, which, again, can potentially result in expensive patent litigation.

The lack of coordination can also lead to research results being generated multiple times. *Shapiro* continues this train of thought and talks about patent thickets defined as "[. . .] a dense web of overlapping intellectual property rights that a company must hack its way through in order to actually commercialize new technology" [41.11]. Many overlapping IPRs might hinder or even prevent new market participants from accessing certain markets, or lead to the fact that companies have to license a large number of property rights from a wide variety of different sources to commercialize their inventions, which may be a very time-consuming and costly undertaking [41.11].

Patent thickets may be formed unintentionally, but are often encouraged by companies acting deliberately. As already mentioned above, because patents can also be used strategically, companies frequently construct patent thickets around their own core inventions in order to preserve their own technological leeway and protect themselves against rivals who *design around* their core patents [41.27, 43]. This problem has been described more generally by *Michael Heller* [41.10, 44] as the *tragedy of the anti-commons*. It can be understood as a dilemma where many rights holders prevent an (optimal) socially desirable outcome [41.10]. Large patent thickets may lead to a product that cannot be produced, since many individually patented components held by different parties increase transaction costs and costs for the necessary licenses [41.10].

An additional problem in the context of patent thickets is the *hold-up problem*, which has also been addressed by *Shapiro* [41.11]. This addresses the issue that new products may violate an existing patent inadvertently, i. e., without knowledge about an existing patent, which may subsequently lead to unanticipated license costs or patent disputes.

As a solution to these problems, *Shapiro* [41.11] suggests cross-licensing, i. e., an exchange of property rights, and patent pools that issue license packages of one or several companies. This does not fully solve the problem, yet would lead to significantly reduced transaction costs [41.9].

41.3 Patentability of Computer Programs—Historical Developments and the Status Quo

Now that we have discussed the pros and cons of patent systems in general, we can turn to the more special discussion in the case of computer-implemented inventions. National patent systems differ strongly with regard to the patentability of software or computer-implemented inventions. This applies, in particular, to the USPTO and the EPO, as well as further national European patent offices.

41.3.1 The United States Patent and Trademark Office

Nowadays, software or more precisely *software-related inventions* are patentable at the USPTO. The patentability of software at the USPTO, however, has a long history, which will be summarized based on the work of *Evans* and *Layne-Farrar* [41.45] and *Cohen* and *Lemley* [41.46].

In the 1970s, software in the USA was viewed as being equivalent to mathematical algorithms or natural laws and was, therefore, not patentable. This changed, however, in 1981 with the United States Supreme Court case *Diamond versus Diehr* [41.47], which can be regarded as setting the precedent for today's software patenting practice in the US. Specifically, this case concerned a molding process for synthetic rubber, where software was used to control and monitor the respective process. The patent was granted based on the argument that a valid patent claim cannot be made invalid by the fact that a computer is involved. This ruling stipulated that the algorithm contained in the software is not protected as an abstract idea in isolation, but exclusively in its application (in this case, the molding of synthetic rubber). Software *as such*, however, was still excluded from patent protection.

This rule remained until a further precedence case occurred in 1994, namely *In re Alappat* [41.48], before the Court of Appeals for the Federal Circuit (CAFC). This case concerned a patent on an invention to display a wave form on a digital oscilloscope. The USPTO had previously declared this patent invalid, because it was not a *special purpose application*, and the same method would also be applicable on a general purpose computer. CAFC, however, regarded the patent as valid, as a general purpose computer would become a special purpose computer if it were programmed to perform specialized functions based on software instructions. Under this ruling, the claims of a patent only had to be designed in a way that a computer program is implemented on any given machine in order to be considered valid.

The remaining obstacle of "being implementable on a machine" was removed in 1995 (*In re Beauregard*, [41.49]), when a patent claim was approved (or was not refused) which protected the computer programs stored on a tangible medium. This meant that in principle software was patentable *as such* at the USPTO. In 1998, the ruling was further softened by the declaration of the invalidity of the *business method exception* in the *State Street versus Signature Financial* case [41.50], in which a *physical structure* was declared not to be necessary if a process or an idea provided a "useful, concrete and tangible result." Thus, and this is the case today, computer programs, software *as such*, and *(internet) business methods* are patentable at the USPTO even without a concrete physical application.

41.3.2 The European Patent Office

EPO was established through the EPC that was signed in Munich in 1973. The EPC came into force in 1977 and the EPO opened its doors in 1978. Since it opened, the EPO has become one of the most important and influential patent offices in the world. As was already stated in Sect. 41.2.1, the European patent system can be regarded as a *system of systems* [41.26]. Alongside European patent law, there are national patent offices that have their own patent law. However, because the national patent systems in Europe are very similar with regard to patenting computer-implemented inventions, the European patent system is described here as representative of the national economies in Europe.

In contrast to the USA, "programs for computers" as well as "plans, rules, and methods [...] for business activities", meaning what has become known as *business methods*, are excluded from patent protection at the EPO (Article 52(2) of the EPC). Yet, this only concerns software as such, which is regulated in Article 52(3) of the EPC. A product or process of a technical nature may be patentable even if the claimed subject matter defines or involves a procedure for a business activity or a computer program [41.51]. The distinguishing feature here is clearly the *technical nature* of an invention. If computer programs perform functions using technical apparatus, are based on technical considerations, have a technical effect, or influence the physical property of an apparatus [41.12] to the extent that a technical nature is given, such an invention is regarded as a computer-implemented invention by the EPO and is, thus, patentable. The exact definition of a computer-implemented invention by the EPO is [41.51]:

> An invention whose implementation involves the use of a computer, computer network, or other programmable apparatus, the invention having one or more features that are realized wholly or partly by means of a computer program is termed a computer-implemented invention.

This is founded on two decisions of the Technical Board of Appeals in 1998 (T935/97 and T1173/97, also known as *computer program product/IBM*) with regard to two patent applications by IBM, which can be regarded as landmark decisions for interpreting Article 52(2) and (3) of the EPC regarding computer programs [41.52]. Here, it was stated that a computer program cannot be excluded from patentability (T 1173/97)

> [...], if it produces a further technical effect which goes beyond the *normal* physical interactions between the program (software) and the computer (hardware)

This implies that software *as such* is still excluded from patentability at the EPO. Yet, there are difficulties with defining the *technical character* or the *technical effect* for patent examiners.

For this reason, in 2002, the European Commission put forward a *Proposal for a Directive of the European Parliament and of the Council on the patentability of computer-implemented* inventions [41.53]. The core of this proposed directive was to make the patenting conditions for computer-implemented inventions more transparent and more specific, but without granting special rights for CII. Instead, the "general patenting conditions should be defined for the specialized requirements of this field of application" [41.54]. Software *as such* and pure business methods should remain excluded from patent protection [41.55]. After several changes in September 2003, however, this proposal was rejected by the European Parliament in July 2005, which is why the decision of the Technical Board of Appeal is still valid at the EPO. This means that computer-implemented inventions can be patented, but only if they have a *technical character* [41.26].

41.4 Definition and Operationalization of Computer-Implemented Inventions

Now that we have a better overview of the recent legal situation with regard to patenting CII, some empirical analyses will provide further information to dig deeper into the economic implications of CII patenting. To this end, however, a technical demarcation of CII for the use within a patent database is necessary. This technical demarcation can be achieved in different ways, for instance, by using technology classifications (e. g., the International Patent Classification (IPC)) or text searches within patent documents.

However, a technical demarcation first of all requires a definition of computer-implemented inventions that can be operationalized. The definition used here is based on the works of *Allison* and *Lemley* [41.56], *Bergstra* and *Klint* [41.57], *Bessen* and *Hunt* [41.58], *Josefsson* [41.59], *Rentocchini* [41.60] and *Xie* and *Miyazaki* [41.61], as well as the definitions used at the *European Patent Office* [41.51] and the *European Commission* [41.53]:

> A computer-implemented invention covers every invention for which a computer, computer network, or other programmable apparatus is used and which has at least one novel characteristic that is realized wholly or partly by means of one or several computer programs. The invention can cover topics related directly to ICT (Information and Communications Technology), e. g., compiling back-ups, data compression, or it can be indirectly related to ICT and only used to control other appliances or devices. Although programs for computers are as such explicitly excluded from patentability (at least at the EPO), a product or a method that is of a technical character, i. e., produces a further (technical) effect beyond the normal functional interaction of program and computer, may be patentable, even if the claimed subject matter defines or at least involves a computer program.

41.4.1 Overview of Already Existing Operationalizations

Economists have already addressed the issue of a technical definition and demarcation of computer-implemented inventions in earlier works. These approaches, however, vary greatly across the respective studies. On the one hand, this is because it is difficult to classify computer-implemented inventions, which is even aggravated by the very technical descriptions of the inventions in patent documents. On the other hand, the objective pursued by the respective authors often differs, which is partly due to the different constraints with regard to patentability at different patent offices, as described above. Economists from the US, for example, often strive to cover the entire software field because software *as such* is patentable at the USPTO. Since this is not the case at the EPO, different definitions apply here.

Graham and *Mowery* [41.62, 63], for example, applied a definition of *software-related inventions* based on technology field classifications for their analyses at the USPTO. In total, 11 IPC-classes (IPC classes G06F 3/*, 5/*, 7/*, 9/*, 11/*, 12/*, 13/*, 15/*; G06K 9/*, 15/*, and H04L 9/*) [41.62] or 12 classes of the U.S. patent classification (USPC) (USPC classes 345, 358, 382, 704, 707, 709–711, 713–715, 717) [41.63], respectively, were applied for their definition. Their assumption was that the entire universe of software patents cannot be mapped exactly over time, but that the IPC or USPC classes used "provide longitudinal coverage of a particularly dynamic and important segment of the overall software industry." To increase the accuracy of their hits, Graham and Mowery additionally limited their analyses to patents of the 100 largest software companies in the USA.

Bessen and *Hunt* [41.58], however, argue in their analysis of software patents at the USPTO, that patent classifications are insufficient to identify software-related inventions, because it is not clearly visible from a patent classification whether the technology is actually a software-related invention. The authors of the study, therefore, use a broad keyword search in the patent specification and description, in which the words *software* or *computer* and *program* must appear.

A combination of several approaches, i. e., a limitation to certain patent classes and keywords appearing in the patent documents is also possible. This was applied by *Allison* and *Tiller* [41.64] in their USPTO study of software patents linked with internet technologies. Similarly, *Schmoch* [41.65] used several procedures—i. e., keywords, a fine-grained IPC class definition, a broader IPC definition, and the keyword *method*—and combined these procedures to identify software patents. He found that especially the keyword *method* led to an increase in the number of software patents. Another possible strategy to identify the relevant patents is to combine search terms with a limitation to companies producing software, as was carried out by *Chabchoub* and *Niosi* [41.66] in a study for American and Canadian companies.

A comparison of these demarcations by *Layne-Farrar* [41.67] (with the exception of *Chabchoub* and *Niosi* [41.66]) shows that *Bessen* and *Hunt* [41.68] identify by far the largest number of patents as *software patents*; the keyword approach, therefore, provides a very broad basis of results. However, *Layne-Farrar* [41.67] was also able to show that the search used by *Graham* and *Mowery* leads to misclassifications of software patents in about 10% of cases, where pure hardware was flagged.

Similarly to *Bessen* and *Hunt* [41.58], *Xie* and *Miyazaki* [41.61] use a keyword search to demarcate the relevant patents at the USPTO in a study of software-related patents in the automobile industry. In contrast to *Bessen* and *Hunt* [41.58], a larger number of keywords is used for the search within titles, abstracts, and patent claims. In addition, for each of the keywords used, *Xie* and *Miyazaki* [41.61] calculated the quality criteria of recall and precision to assess the accuracy of each individual keyword.

41.4.2 The Demarcation of CII

As a basis for the operationalization of CII in this paper, we draw on the work of *Xie* and *Miyazaki* [41.61], who evaluated the effectiveness of keyword search strategies for the identification of CII patents, employing a sample of embedded software-related patents in the domain of automobiles.

As a first step, we selected all the keywords with a precision value above 90% from the keyword list provided by *Xie* and *Miyazaki* [41.61] (Table 41.1). The precision measures the share of correctly identified elements in all identified elements, whereas the recall measures the share of all correctly identified elements in the total number of relevant elements. Precision, thus, indicates the accuracy of a method, while recall indicates its yield. Generally, an increase in precision implies a lower number of errors but also leads to lower recall rates and vice versa. This high-precision approach of *Xie* and *Miyazaki* [41.61] reduces the overall number of hits, but also means that there is a higher probability of only capturing patents that actually protect computer-implemented inventions. We searched for these keywords in the titles, abstracts, and claims patents of all patent filings at the EPO. Unfortunately, we did not have access to the patent claim data at the USPTO, which is why we could only search titles and abstracts here.

In a second step, the patent applications identified via the keyword search were crossed with the WIPO list of 35 technology fields [41.69]), in order to be able to calculate the shares of identified patents in total patent applications of the respective technology field. In computer technology and data processing, 74% of all filings were identified with our keyword search. This makes it clear that, as already indicated by *Layne-*

Table 41.1 List of keywords

Keyword	Recall	Precision
[Micro]processor	18.6	100
Chip	0.7	100
Comput% program	8.8	100
Controller	26.0	100
Data	31.9	100
Digital	7.8	100
Integrated circuit	2.0	100
Image processing	1.7	100
Information processing	0.5	100
Processing unit	3.7	100
Program%	13.7	100
Software	5.4	100
Comput%	28.2	99.1
Signal processing	15.0	98.4
Identify%	10.0	97.6
Control unit	15.2	95.4
Memory	15.9	94.2
Calculat%	19.6	94.1
Electronic%	18.1	93.7
Monitoring	10.3	93.3
Imaging	2.9	92.3

Source: Own presentation based on *Xie* and *Miyazaki* [41.61]. Note: Recall and precision are illustrated here when using the keywords in titles, abstracts, and patent claims following *Xie* and *Miyazaki* [41.61]. For the analyses here, only keywords with more than 90% precision were used. In addition, the keyword *information* was excluded from the demarcation because of too many Type II errors

Farrar [41.67], patents filed in the field of computer technology and data processing can partly be classified as pure hardware. Across all the other technology fields, the share of patents identified using keywords is much smaller. However, it is apparent that computer-implemented inventions can be found across the entire range of technology fields. Besides the fields related to electrical engineering, where the shares of CII patents are mostly above the 50% mark, there are comparatively high shares in medical technology (24%) and in mechanical engineering and automobile manufacturing (between 10 and 18%). Even in fields related to chemistry, shares of 4 to 6% were found. The scattering of patents across all fields indicates that any limitation to specific technology fields would result in a large number of CII patents being unidentified. Yet, a demarcation based purely on IPC classes would probably produce a large number of irrelevant hits, as discussed by *Layne-Farrar* [41.67].

In the third step, the full texts of ten patents per technology field tagged as CII by our keyword search were examined manually by experts in the field. It was found that none of the patents identified in the field *pharmaceutical products* were actually intended to protect a computer-implemented invention. To eliminate this erroneous classification, all the patents that belong exclusively to this technology field were excluded from the analysis, i. e., they were tagged as *pure hardware*.

In a final step, the distribution of the identified CII patents was calculated according to the classes of the IPC at the four-digit level. To a large extent, the results mirror the picture resulting from the analysis of the technology fields, i. e., we find a large spread of CII patents across technologies. During the manual identification, however, several filings that were suspected of protecting software *as such* (these were not necessarily granted) were found. In order to rule out that patents for software *as such* appear in our analysis, we excluded the patent classes H04L 29/06, G06F 11/30, G06F 17/24, G06F 17/30, G06Q 10, G06F 9/00, G06F 9/06, G06F 9/2, G06 9/3, G06F 9/4, and G06F 9/5 (including existing sub-classes) in case these were the only classes stated on a given patent filing. This does not imply that all these patents actually concern software *as such*, but only that including them would generate a larger fuzziness of hits. To this extent, our demarcation represents a conservative estimate, and its results in terms of numbers should be positioned at the lower end of the *real* distribution.

However, even with this stepwise procedure, we cannot rule out that our classification is incorrect for a certain share of patents, although the conservative approach should ensure that software *as such* is excluded to the greatest extent possible. As already argued by *Graham* and *Mowery* [41.62], the entire universe of patents for CII cannot be exactly identified, even though we have tried to eliminate potential sources of errors as far as possible.

41.4.3 The Database

The patent data for were extracted from the *EPO Worldwide Patent Statistical Database* (PATSTAT), which provides information about published patents from more than 80 patent authorities worldwide. We restricted our analyses to filings at the EPO and the USPTO. All the filings were counted using the year of the worldwide first filing, i. e., the priority year. This is the earliest registered date in the patent process and is, therefore, closest to the date of invention. For further analyses, we also introduced a distinction by the type of patent applicant, i. e., whether it is a large or small or medium sized enterprise (SME). In order to identify SMEs within the database, we matched PATSTAT with *Bureau van Dijk's* global company database *ORBIS*, where information on the number of employees is available. The matching routine is based on a probability matching (Levenshtein distance) of the name of the patent applicant in PATSTAT and the company name in ORBIS. For the identification of companies, we first of all excluded single inventors by comparing the names of the inventor with the name of the patent applicant. Subsequently, universities and other research organizations were manually coded, which resulted in a dataset featuring companies only. All companies with less than 500 employees were then classified as SMEs, and all others were tagged as large firms. Companies with no available employee information (or the ones that could not be identified during matching to ORBIS) were classified using their patent numbers, i. e., all the companies with less than 10 filing applications in the period 2000–2011 were coded as SMEs. In a final step, the lists of SMEs and large enterprises were manually checked and corrected. Besides the information on employees, the ORBIS database allows us to find out the sector classification (NACE codes—Nomenclature Statistique des Activités Énomiques dans la Communauté Européenne/Statistical Classification of Economic Activities in the European Community) of the respective companies. This enables us to analyze and calculate the spread of CII filings across economic sectors.

41.5 Empirical Trends in CII Filings

In this section, we will provide some empirical trends in CII patenting. This is first of all on the level and amount of CII patents to provide some evidence on the magnitude of CII in the U.S. and Europe. Apart from that, we will provide some further rather basic empirical facts about the enrollment of SMEs within the field and the spread of the technology across economic sectors.

The absolute numbers of CII filings at the USPTO and the EPO are depicted in Fig. 41.1. In 2012, about 120 000 CII patents were filed at the USPTO. There was an obvious growth in CII filings between 2001 and 2005, before the figures started to decline during the economic crisis of 2008 and 2009, which, however, is a trend that is obvious across all technology fields. In 2010, the figures started to climb again to reach their peak in 2012.

Similar trends are visible at the EPO, albeit at a lower level overall. It is interesting to note that although the trends are similar in relative terms, the search of keywords in patent claims leads to a much

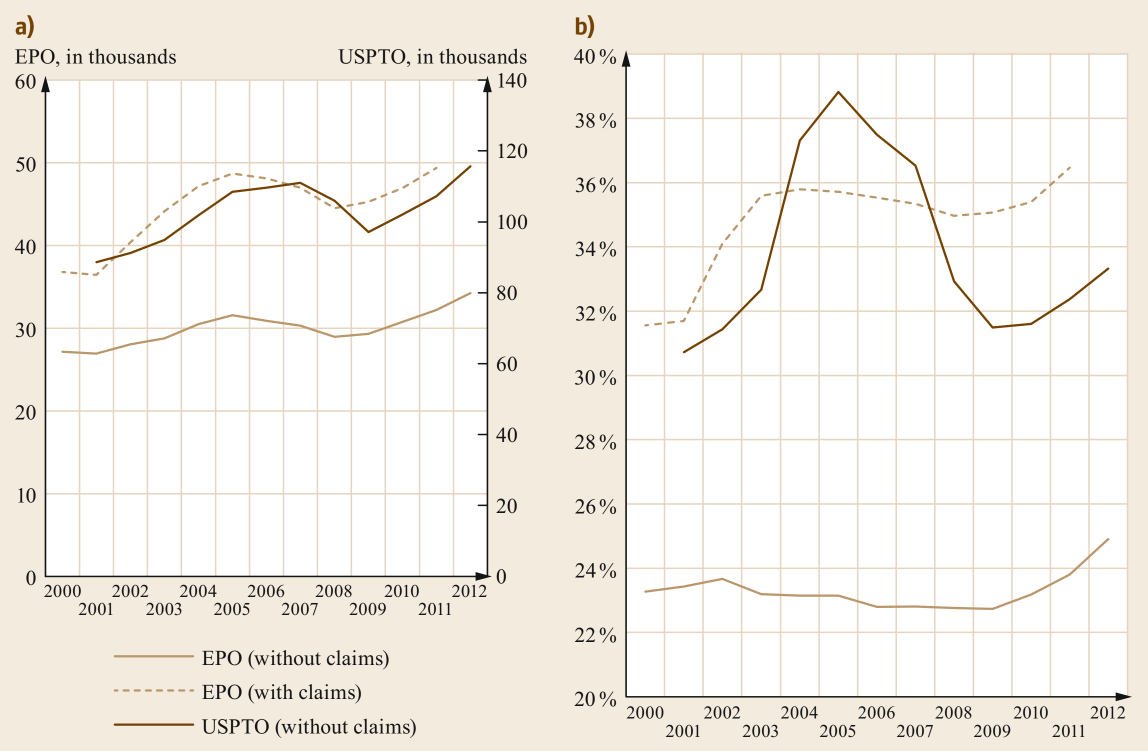


Fig. 41.1 **(a)** Absolute number and **(b)** shares of CII applications at the EPO and the USPTO. Source: EPO-PATSTAT, own calculations

larger absolute number of identified CII filings, implying that a) neglecting patent claims in keyword searches, especially in CII, delivers quite conservative, lower-bound estimates for the true amount of filings, and b) the *real* USPTO figures probably are larger than the ones identified in our search. If the search is limited to the titles and abstracts of the patent filings, the number of CII filings at the EPO amounts to a current margin of approx 34 000. If patent claims are included in the search, this figure increases to about 50 000 (in 2011).

When looking at the shares of CII filings in all filings (at the respective patent office), it can be found that about 32 to 39% of all applications at the USPTO can be classified as CII patents. Yet, the development is quite dynamic. While the shares increased steeply between 2001 and 2005, it declined almost as quickly again afterwards. After the economic crisis in 2009, however, we once again observe a rise in CII filings. The situation at the EPO, however, varies depending on the definition used. If patent claims are included in the search, a rising trend emerges that is similar to that in the US at the beginning of the millennium. However, the decline at the EPO is much less dramatic. If the search is restricted to titles and abstracts, it becomes apparent that the share of CII filings remains more or less constant over time at around 23% of all EPO filings until 2009, where an increase in filings can be observed. This indicates two things. The first one is that especially at the beginning of the 2000s, fewer IT-relevant keywords in the titles and abstracts of patents were used, and the claims delivered more detailed descriptions. The second one is that the rise in the shares of CII filings occurred after the crisis of the *new economy* and at the same time as the above mentioned European Commission's *Proposal for a Directive of the European Parliament and of the Council for the patentability of computer-implemented inventions*. Unfortunately, we do not have the data to control whether there are causal relationships between these events and the rise in CII shares.

Overall, however, it can be concluded that almost one-third (or more conservatively, i. e., without including claims, approximately one-quarter) of all patent filings at the EPO and at the USPTO are CII applications. We, therefore, do not target a marginal phenomenon but discuss a major share of patent filings at the respective offices. These figures on their own are evidence that clarification is needed with regard to the definition and demarcation of the *technical character* or *technical effect* of an invention at the EPO in order to take account of the factual relevance of such inventions. The patent system should provide clear rules for applicants (and potential applicants to decrease uncertainty and related costs).

This issue is especially relevant for SMEs. From Fig. 41.2, it becomes apparent that SMEs are underrepresented in the field of CII at the EPO and the USPTO. In other words, the share of SME filings on average is larger in total patent filings than it is in the field of CII. While the total share of SME filings at the USPTO was about 25% in 2012, SME shares in CII only reached 21%. For the EPO, the effect is even more strongly pronounced. Here, the average share of SME filings reached 29% in 2012, while the share of SME filings in CII was 23% (or 22% in the case when patent claims are used in the search).

Besides the dimension of size, the sectoral distribution (NACE Rev. 2, 2-digit) of CII patent filings offers interesting insights (Fig. 41.3). It can be found that CII patents are spread widely across economic sectors. Not only firms from the electrical industry, i. e., *computer, electronic, and optical products* and *electrical equipment* file CII patents. In fact, approximately 41% of all CII applications within manufacturing at the EPO orig-

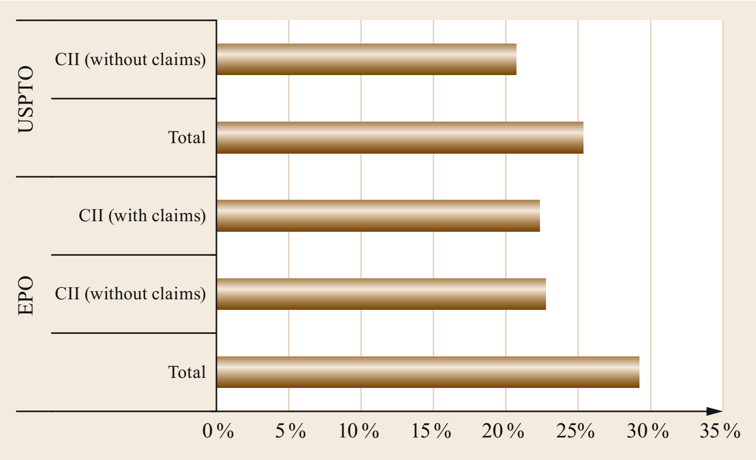


Fig. 41.2 Shares of SME filings in all filings from industry, 2012. Source: EPO-PATSTAT, BvD ORBIS, own calculations

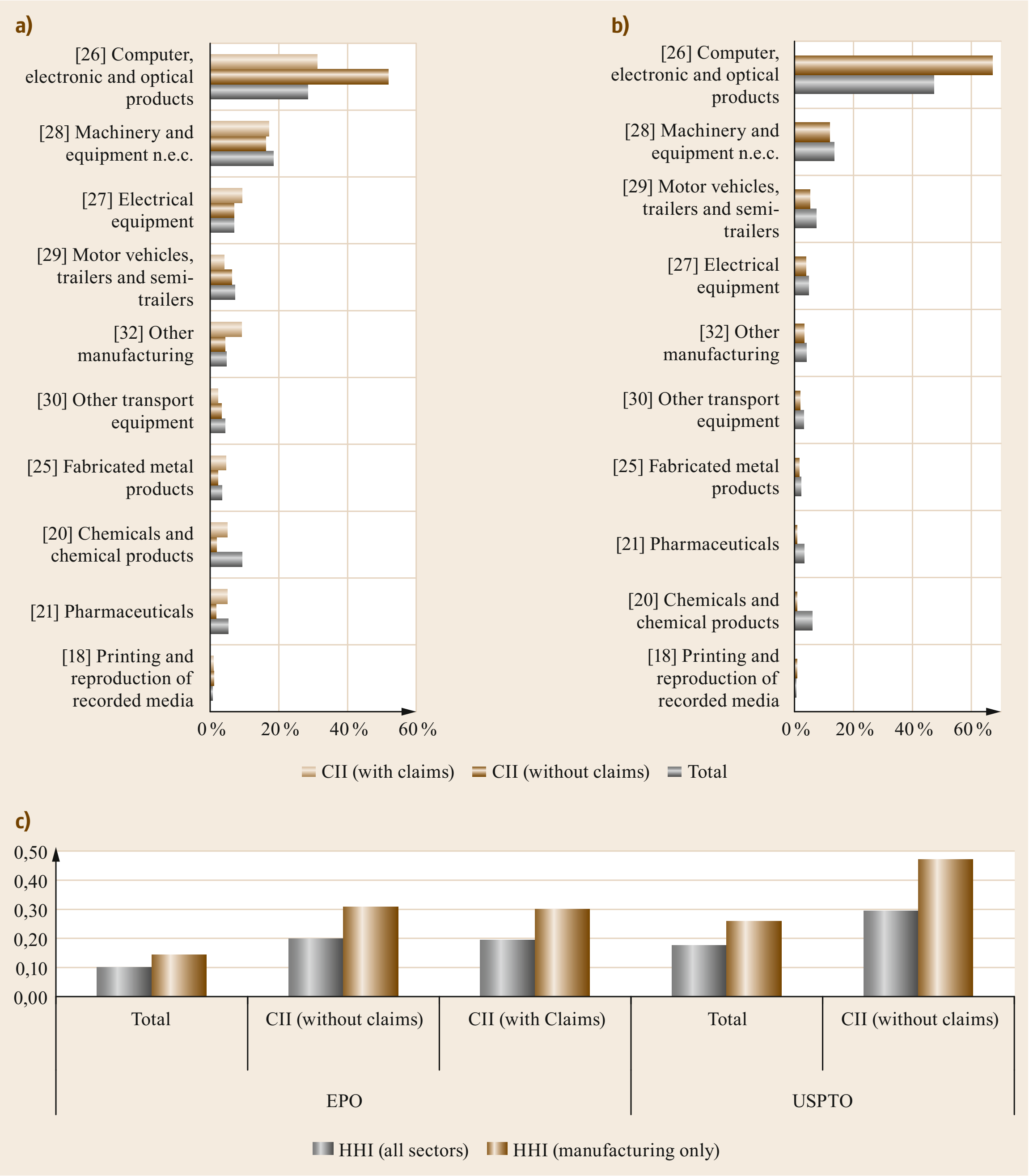


Fig. 41.3a–c Sectoral distribution of CII filings within manufacturing, 2010–2012. **(a)** EPO, **(b)** USPTO, **(c)** Herfindahl-Hirschman index. Source: EPO-PATSTAT, BvD ORBIS, own calculations. Only sectors with CII shares above 1% are shown. n.e.c. – Not elsewhere classified

inate from other industries (Fig. 41.3a). This further depends on the search strategy and the patent office under analysis. With claims in the search included, this share even rises to nearly 60%. At the USPTO (Fig. 41.3b), the share is somewhat smaller, with 33%. It can, thus, be seen as a cross-cutting or *general purpose* technology, where technology is not only used but also developed in other sectors. This is further backed by the Herfindahl–Hirschman index (HHI) calculations provided in the lower panel of the Fig. 41.3c. The HHI is a concentration measure ranging from 0 (equal distribution) to 1 (full concentration). Here, relatively small concentration values can be observed. The main non-electronics related sector where CII filings are produced is *machinery and equipment* which, with 17%, has a high proportion of CII patents. Nearly every fifth CII patent within manufacturing is thus filed by a company from the machinery sector. This share even becomes larger when taking into account the full-text of patents in the keyword search. Consequently, we can talk about a computerization of mechanical engineering, which is often summarized under the catchphrase *digitalization* or *Industry 4.0*. Computer technologies are not just applied here but also contribute to technological advance. Relevant shares of CII patents, however, are also found in the chemical industry (chemicals and pharmaceuticals), especially at the EPO. CII patents are, therefore, also filed by companies in economic sectors that, at least at first sight, have nothing to do with computer technology.

Finally, it is interesting to take a closer look at the legal status of CII filings (Fig. 41.4), especially against the background that software as such is not patentable at the EPO (this analysis could only be performed for the EPO, as the USPTO data was not available). First of all, it was found that CII patents are withdrawn more frequently than the average patent, although also the average withdrawal rate has increased since 2005. This might have two reasons. First, computers and related technologies mature rather quickly as technological progress in the field has been quite fast paced, especially in the past 20 years. It is, thus, a reasonable strategy for patent applicants to withdraw their filings before grant in case there are new technological developments to avoid further procedural costs at the patent office.

A second reason could be that signals from the patent office, i. e., preliminary search reports, communications, etc., which provide the applicant with the information that his or her patent only has a low chance of being granted (anticipated refusal), might lead to an early withdrawal. This is also reflected in the lower grant rate of CII filings, which, however, converged to a certain extent with the average grant rate in 2009/2010. With the data at hand, we are not able to test whether this has to do with a certain share of software *as such* that is filed at the EPO. Yet, the higher withdrawal and smaller grant rate at least point in this direction.

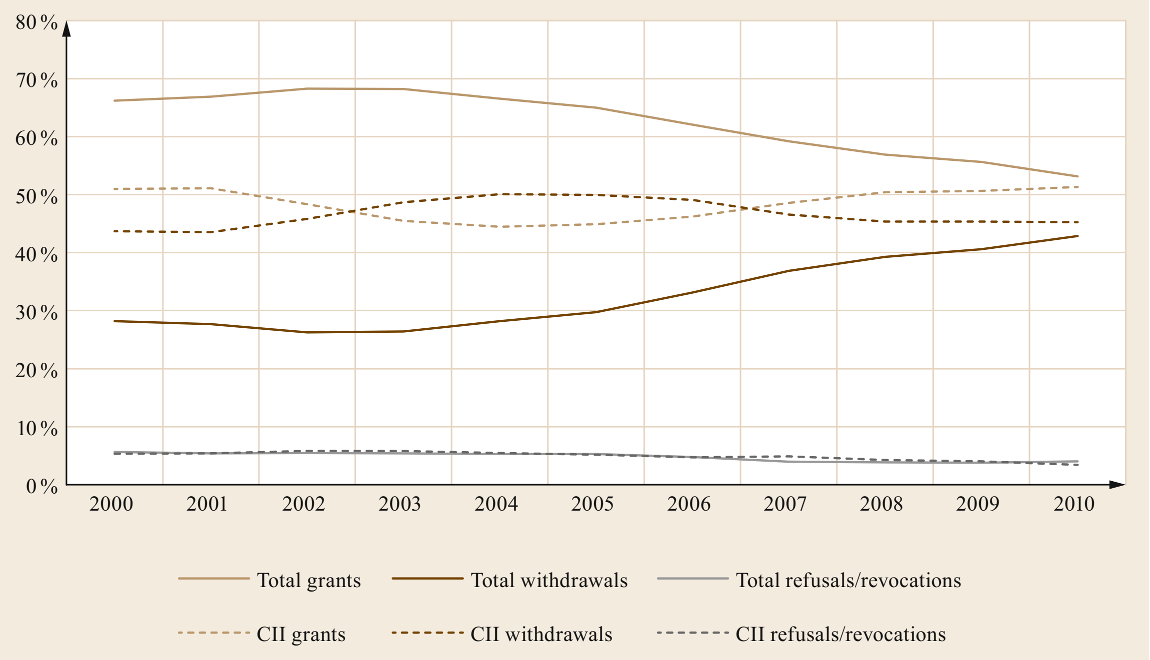


Fig. 41.4 Legal status of CII filings at the EPO. Source: EPO-PATSTAT, own calculations

41.6 Summary and Implications

With regard to patenting computer programs there are still large differences in the European and the U.S. patent system. While software *as such* can be patented at the USPTO, the EPO prohibits patenting pure computer programs. The distinguishing feature for the EPO is the *technical character* of an invention. A computer program is only patentable if it is of *technical nature* or has a *technical effect* that goes beyond the *normal* physical interactions between a program (software) and a computer (hardware). A product or process of a technical nature may, thus, be patentable even if a computer program is included.

These differences are, of course, reflected in the numbers of CII patents being filed. We applied a conservative estimate, which places the share of CII filings at the EPO at around 25% at present, while at the USPTO a current margin of approximately 33% can be reached. Unfortunately, data availability limits this analysis to the EPO. However, using the EPO figures as a benchmark, we would expect even higher shares for the USPTO when including the patent claims in the search. It can, therefore, be concluded that at least every fourth patent at the EPO and every third patent at the USPTO is, in fact, a CII filing, i. e., we are indeed talking about a major share of filings at the respective offices.

In addition, we found that CII patents are filed more often by large enterprises than by SMEs. In comparison to the total patents, CII shares are over-represented in the portfolios of large firms. CII patents are also widely spread across economic sectors. Not only firms from the electrical industry file CII patents. Computer technologies are to a large extent not only used, but also produced by firms from other sectors, mostly within the machinery industry. However, we should not focus solely on the *computerization* or *digitalization* of machinery. Patents for computer technology can, to a greater or lesser extent, be found in almost every industry of the manufacturing sector, among them also the chemical and pharmaceutical industry.

These trends provide evidence that clarification is needed with regard to the definition and demarcation of the *technical character* or *technical effect* of an invention at the EPO, in order to take account of the factual relevance of such inventions. Clear rules are essential to reduce uncertainties and provide the relevant incentives for innovation. With regard to patenting computer programs, these seem to be at least partly lacking at the moment. This weakens the patent system, which is designed as an incentive mechanism to generate and diffuse knowledge and increase innovative capacity. Yet, the benefits of the system need to exceed its costs. Increased uncertainty, however, is a cost factor for patenting firms, i. e., for information search, etc., which might lead to *social* costs through the emergence of patent thickets that can lead to an increase in costly patent disputes. Overlapping patent rights can also block (further) technological developments, especially for new and complex technologies, which manifests itself, for example, in high market entry barriers for innovative companies. For products that consist of many single components, overlapping patent rights also raise transaction costs, e. g., due to high licensing costs or the (too) high risk of a patent lawsuit. This may also be a reason why especially small companies, which tend to lack the relevant resources, file relatively fewer patents for CII.

Overall, and especially given the backdrop of lower patent shares of SME—it must be ensured that planning security in the patent system is guaranteed by clear and transparent rules. This starts by defining the *technical character* of an invention at the EPO. Taking this argument a step further, it would be desirable to have uniform rules for CII patenting worldwide to reduce transaction costs and ensure an innovation fostering patent system.

References

41.1 H. Grupp: *Foundations of the Economics of Innovation–Theory, Measurement and Practice* (Edward Elgar, Cheltenham 1998)

41.2 S.R. Adams: *Information Sources in Patents* (K.G. Saur, Munich 2006)

41.3 R. Frietsch, U. Schmoch, B. van Looy, J.P. Walsh, R. Devroede, M. Du Plessis, T. Jung, Y. Meng, P. Neuhäusler, B. Peeters, T. Schubert: *The Value and Indicator Function of Patents*, Studien zum deutschen Innovationssystem, Vol. 15-2010 (Expertenkommission Forschung und Innovation, Berlin 2010)

41.4 R. Frietsch, P. Neuhäusler, O. Rothengatter: *Patent Applications–Structures, Trends and Recent Developments*, Studien zum deutschen Innovationssystem, Vol. 8-2012 (Expertenkommission Forschung und Innovation, Berlin 2012)

41.5 U. Schmoch, H. Grupp: *Wettbewerbsvorsprung durch Patentinformation – Handbuch für die Recherchepraxis. Karlsruhe, Fraunhofer-Institut für*

Systemtechnik und Innovationsforschung (ISI) (Verlag TÜV Rheinland, Köln 1990)
41.6 J. Bessen, M.J. Meurer: *Patent Failure: How Judges, Bureaucrats, and Lawyers Put Innovators at Risk* (Princeton Univ. Press, Princeton 2008)
41.7 J. Bessen, M.J. Meurer: *The Patent Litigation Explosion*, Working Paper Series Law and Economics, Vol. 05–18 (Univ. School of Law, Boston, Boston 2005)
41.8 J. Bessen, M.J. Meurer: What's wrong with the patent system? Fuzzy boundaries and the patent tax, First Monday (2007), https://doi.org/10.5210/fm.v12i6.1867
41.9 R.W. Hahn: An overview of the economics of intellectual property protection. In: *Intellectual Property Rights in Frontier Industries: Software and Biotechnology*, ed. by R.W. Hahn (AEI-Brookings Joint Center for Regulatory Studies, Washington 2005) pp. 11–44
41.10 M.A. Heller, R.S. Eisenberg: Can patents deter innovation? The anticommons in biomedical research, Science (1998), https://doi.org/10.1126/science.280.5364.698
41.11 C. Shapiro: Navigating the patent thicket: Cross licenses, patent pools, and standard-setting. In: *Innovation Policy and the Economy*, NBER Book Series Innovation Policy and the Economy, Vol. 1, ed. by B.A. Jaffe, J. Lerner, S. Stern (MIT Press, Cambridge 2001) pp. 119–150
41.12 K. Blind, J. Edler, R. Nack, J. Straus, M. Friedewald, R. Frietsch (Eds.): *Software-Patenten. Eine empirische Analyse aus ökonomischer und juristischer Perspektive*, Schriftenreihe des Fraunhofer-Instituts für Systemtechnik und Innovationsforschung ISI "Technology, Innovation and Policy", Vol. 49 (Physica, Heidelberg 2002)
41.13 K. Blind, J. Edler, M. Friedewald: *Software Patents, an Empirical Analysis from an Economic Perspective* (Fraunhofer IRB, Stuttgart 2004)
41.14 K. Blind, J. Edler, M. Friedewald: *Software Patents, Economic Impacts and Policy Implications, New horizons in intellectual property* (Elgar, Cheltenham 2005)
41.15 C. Rammer: *Innovationsverhalten der Unternehmen in Deutschland 2005 - Aktuelle Entwicklungen – öffentliche Förderung – Innovationskooperationen – Schutzmaßnahmen für geistiges Eigentum*, Studien zum deutschen Innovationssystem, Vol. 13-2007 (Zentrum für Europäische Wirtschaftsforschung (ZEW), Mannheim 2007)
41.16 C. Edquist, M. McKelvey: *Systems of Innovation, Growth, Competitiveness and Employment* (Edward Elgar, Northampton 2000)
41.17 B.-A. Lundvall, D. Foray: The knowledge-based economy. In: *Employment and Growth in the Knowledge-Based Economy*, ed. by OECD (OECD, Paris 1996)
41.18 E. Malecki: *Technology and Economic Development. The Dynamics of Local, Regional and National Change* (Longman, Harlow 1991)
41.19 R.R. Nelson, P.M. Romer: Science, economic growth, and public policy. In: *Technology, R&D, and the Economy*, ed. by B.L.R. Smith, C.E. Barfield (The Brookings Institution, Washington 1996)
41.20 P.M. Romer: The origins of endogenous growth, J. Econ. Perspect. (1994), https://doi.org/10.1257/jep.8.1.3
41.21 P. Hanel: The use of intellectual property rights and innovation by manufacturing firms in Canada, Econ. Innov. New Technol. (2008), https://doi.org/10.1080/10438590701581481
41.22 J.E. Stiglitz: *The Theory of International Public Goods and the Architecture of International Organizations*, United Nations Background Paper, Vol. 7 (United Nations, Department for Economic and Social Information and Policy Analysis, New York 1995)
41.23 J.E. Stiglitz: Economic foundations of intellectual property rights, Duke Law J. **57**(6), 1693–1724 (2008)
41.24 K. Arrow: Economic welfare and the allocation of resources for invention. In: *The Rate and Direction of Inventive Activity: Economic and Social Factors, Economic Welfare and the Allocation of Resources for Invention*, ed. by R. Nelson (Princeton Univ. Press, Princeton 1962) pp. 609–626
41.25 J.E. Stiglitz: Knowledge as a global public good. In: *Global Public Goods: International Cooperation in the 21st Century*, ed. by I. Kaul, I. Grunberg, M. Stern (Oxford Univ. Press, New York 1999)
41.26 S. Borrás, B. Kahin: Patent reform in Europe and the US, Sci. Public Policy **36**(8), 631–640 (2009), https://doi.org/10.3152/030234209X474589
41.27 K. Blind, J. Edler, R. Frietsch, U. Schmoch: Motives to patent: Empirical evidence from Germany, Res. Policy **35**(5), 655–672 (2006), https://doi.org/10.1016/j.respol.2006.03.002
41.28 K. Blind, J. Edler, R. Frietsch, U. Schmoch: *Erfindungen kontra Patenten. Schwerpunktstudie "zur technologischen Leistungsfähigkeit Deutschlands", Endbericht für das Bundesministerium für Bildung und Forschung (BMBF)* (Fraunhofer-Institut für System- und Innovationsforschung, Karlsruhe 2003)
41.29 W.M. Cohen, R.R. Nelson, J.P. Walsh: *Appropriability Conditions and why Firms Patent and why They Do not*, Working Paper, Vol. 7552 (National Bureau of Economic Research (NBER), Cambridge 2000)
41.30 N. Harabi: Appropriability of technical innovations: An empirical analysis, Res. Policy (1995), https://doi.org/10.1016/0048-7333(94)00812-4
41.31 P. Neuhäusler: The use of patents and informal appropriation mechanisms—Differences between sectors and among companies, Technovation (2012), https://doi.org/10.1016/j.technovation.2012.07.004
41.32 A. Arundel, P. Patel: *Strategic Patenting – Background Report for the Trend Chart Policy Benchmarking Workshop "New Trends in IPR Policy"* (2003)
41.33 D. Guellec: Patents as an incentive to innovate. In: *The Economics of the European Patent System—IP Policy for Innovation and Competition*, ed.

by D. Guellec, B. van Pottelsberghe (Oxford Univ. Press, Oxford 2007) pp. 46–84
41.34 C. Mersch: *Die Welt der Patente – Soziologische Perspektiven auf eine zentrale Institution der globalen Wissensgesellschaft* (transcript, Bielefeld 2013)
41.35 E. Kitch: The nature and function of the patent system, J. Law Econ. (1977), https://doi.org/10.1086/466903
41.36 N.T. Gallini, R.A. Winter: Licensing in the theory of innovation, Rand J. Econ. (1985), https://doi.org/10.2307/2555412
41.37 Z. Griliches: The search for R&D spillovers, Scand. J. Econ. (1992), https://doi.org/10.3386/w3768
41.38 A.B. Jaffe, M. Trajtenberg, R. Henderson: Geographic localization of knowledge spillovers as evidenced by patent citations, Q. J. Econ. (1993), https://doi.org/10.2307/2118401
41.39 A.B. Jaffe: Technological opportunity and spillovers of R&D: evidence from firm's patents, profits, and market value, Am. Econ. Rev. (1986), https://doi.org/10.3386/w1815
41.40 A.B. Jaffe, M.S. Fogarty, B.A. Banks: Evidence from patents and patent citations on the impact of NASA and other federal labs on commercial innovation, J. Ind. Econ. (1998), https://doi.org/10.3386/w6044
41.41 A.B. Jaffe, M. Trajtenberg, M.S. Fogarty: Knowledge spillovers and patent citations: evidence from a survey of inventors, Am. Econ. Rev. (2000), https://doi.org/10.1257/aer.90.2.215
41.42 N. Thumm: *Research and Patenting in Biotechnology: A Survey in Switzerland* (Swiss Federal Institute of Intellectual Property, Bern 2003)
41.43 D.L. Rubinfeld, R. Maness: The strategic use of patents: Implications for antitrust. In: *Antitrust, Patents and Copyright: EU and US Perspectives*, ed. by F. Leveque, H. Shelanski (Edward Elgar, Northampton 2005) pp. 85–102
41.44 M.A. Heller: The tragedy of the anticommons: Property in the transition from Marx to markets, Harv. Law Rev. **111**(3), 621–688 (1998)
41.45 D.S. Evans, A. Layne-Farrar: Software patents and open source: The battle over intellectual property rights, Va. J. Law Technol. **10**(10) (2004)
41.46 J.E. Cohen, M.A. Lemley: Patent scope and innovation in the software industry, Calif. Law Rev. (2001), https://doi.org/10.15779/Z38ZD8J
41.47 Supreme Court of the United States: Diamond vs. Diehr: 450 U.S. 175, 185 (1981), Full case name: Diamond, Commissioner of Patents and Trademarks v. Diehr, et al.
41.48 United States Court of Appeals for the Federal Circuit: In re Alappat: 33 F.3d 1526, 1537 (Fed. Cir. 1994), Full case name: In re Kuriappan P. Alappat, Edward E. Averill and James G. Larsen
41.49 United States Court of Appeals for the Federal Circuit: In re Beauregard: 53 F.3d 1583, 1584 (Fed. Cir. 1995)
41.50 United States Court of Appeals for the Federal Circuit: State Street vs. Signature Financial: 149 F.3d 1368, 1370 (Fed. Cir. 1998), Full case name: STATE STREET BANK & TRUST CO., Plaintiff-Appellee, v. SIGNATURE FINANCIAL GROUP, INC., Defendant-Appellant
41.51 European Patent Office: Examination of computer-implemented inventions at the European Patent Office with particular attention to computer-implemented business methods, Official J. EPO **11**, 594–600 (2007)
41.52 I. Muir, M. Brandi-Dohrn, S. Gruber: *European Patent Law: Law and Procedure Under the EPC and PCT* (Oxford Univ. Press, Oxford 2002)
41.53 European Commission: Proposal for a directive of the European Parliament and of the Council on the patentability of computer-implemented inventions, Official J. Eur. Communities COM/2002/0092 (2002)
41.54 Federal Court of Justice (Bundesgerichtshof – BGH): Patentierbarkeit computerimplementierter Erfindungen, http://www.gesmat.bundesgerichtshof.de/gesetzesmaterialien/15_wp/computerimplerfind/compimplerf-index.htm (2008)
41.55 Federal Ministry of Justice and Consumer Protection (Bundesministerium der Justiz – BMJ): *Press Release: EP weist Standpunkt zu computerimplementierten Erfindungen zurück* (Referat Presse- und Öffentlichkeitsarbeit des Bundesministeriums der Justiz, Berlin 2005)
41.56 J.R. Allison, M.A. Lemley: Who's patenting what? An empirical exploration of patent prosecution, Vanderbilt Law Rev. **53**, 2099 (2000), https://doi.org/10.2139/ssrn.223312
41.57 J.A. Bergstra, P. Klint: How to find a software patent? In: *Draft Interim Deliverable for an EC Consulting Project on Software Patents* (2007)
41.58 J. Bessen, R.M. Hunt: An empirical look at software patents, J. Econ. Manag. Strategy **16**(1), 157–189 (2007), https://doi.org/10.1111/j.1530-9134.2007.00136.x
41.59 E. Josefsson: The Gauss project: foundation for a free information infrastructure. In: *5th EPIP Conf.: Eur. Policy Patents Intellectual Property: What Direction Should it Go?* (EPIP, Maastricht 2005)
41.60 F. Rentocchini: Sources and characteristics of software patents in the European Union: Some empirical considerations, Inf. Econ. Policy (2011), https://doi.org/10.1016/j.infoecopol.2010.12.002
41.61 Z. Xie, K. Miyazaki: Evaluating the effectiveness of keyword search strategy for patent identification, World Patent Inf. (2013), https://doi.org/10.1016/j.wpi.2012.10.005
41.62 S.J.H. Graham, D.C. Mowery: Intellectual property protection in the U.S. software industry. In: *Patents in the Knowledge-Based Economy*, ed. by W.M. Cohen, S.A. Merrill (National Academies, Washington 2003)
41.63 S.J.H. Graham, D.C. Mowery: Software patents: Good news or bad news? In: *Intellectual Property Rights in Frontier Industries: Software and Biotechnology*, ed. by R.W. Hahn (AEI-Brookings Joint Center for Regulatory Studies, Washington 2005) pp. 45–80

41.64 J.R. Allison, E.H. Tiller: Internet business method patents. In: *Patents in the Knowledge-Based Economy*, ed. by W.M. Cohen, S.A. Merrill (National Academies, Washington 2003)

41.65 U. Schmoch: *Definition of Patent Search Strategies for Selected Technological Areas* (Fraunhofer ISI, Karlsruhe 2003), Report to the OECD

41.66 N. Chabchoub, J. Niosi: Explaining the propensity to patent computer software, Technovation **25**(9), 971–978 (2005), https://doi.org/10.1016/j.technovation.2004.02.015

41.67 A. Layne-Farrar: *Defining Software Patents: A Research Field Guide*, AEI-Brookings Joint Center Working Paper, Vol. 05-14 (AEI-Brookings Joint Center for Regulatory Studies, Washington 2005)

41.68 J. Bessen, R.M. Hunt: *An Empirical Look at Software Patents*, Working Paper, Vol. 03-17/R (Federal Reserve Bank of Philadelphia, Philadelphia 2004)

41.69 U. Schmoch: *Concept of a Technology Classification for Country Comparisons* (Fraunhofer ISI, Karlsruhe 2008), Final Report to the World Intellectual Property Office (WIPO)

Peter Neuhäusler

Competence Center Policy – Industry – Innovation
Fraunhofer Institute for Systems and Innovation Research ISI
Karlsruhe, Germany
peter.neuhaeusler@isi.fraunhofer.de

Peter Neuhäusler is a Project Manager at the Competence Center Policy – Industry – Innovation at the Fraunhofer Institute for Systems and Innovation Research (Karlsruhe). He studied Social Sciences, majoring in Micro- and Macro-Sociology, Applied Social Research, and Quantitative Methodology at the University of Mannheim and graduated with a diploma in 2008. Peter Neuhäusler was awarded his Doctoral degree in Innovation Economics at the Technical University of Berlin in 2012.

Rainer Frietsch

Competence Center Policy – Industry – Innovation
Fraunhofer Institute for Systems and Innovation Research ISI
Karlsruhe, Germany
rainer.frietsch@isi.fraunhofer.de

Rainer Frietsch is Deputy Head of the Competence Center Policy – Industry – Innovation and Coordinator of the Business Unit Innovation Indicators at the Fraunhofer Institute for Systems and Innovation Research ISI. Rainer completed business training in wholesale and import/export trade. After that he studied Social Science at the University of Mannheim. He holds a PhD in Economics from the Karlsruhe Institute of Technology (KIT).

42. Interplay of Patents and Trademarks as Tools in Economic Competition

Sandro Mendonça, Ulrich Schmoch, Peter Neuhäusler

Integrated manufacturing-service systems have been receiving attention recently. The phenomenon of services-to-artifacts companies, namely those specializing in intermediate goods and complex equipment, is increasingly instrumental for long-run competitiveness in fast-changing, high-quality global markets. The debate has so far has remained largely qualitative, and the effective role and relevance of services is rather fuzzy. Against this background, this chapter brings in empirical evidence concerning the evolving business models of a variety of leading innovative manufacturing companies. For this purpose, over 50 manufacturing companies listed in the European Union (EU) research & development (R&D) investment scoreboard are analyzed in terms of patents and trademarks. In particular, trademark strategies are studied in greater depth, and they are sub-divided into *goods* and *services* marks and into *high* and *low* sophistication. Service marks are used as a supplement to patents, as the service component of industrial offerings is not covered by classic indicators of technical change. The economic data from the EU (EU Scoreboard R&D, sales, growth, employees, profits, or investment) are linked to the patent and trademark data in order to see which balance of goods and service capabilities leads to favorable economic results.

42.1 Pattern of R&D-Intensive Enterprises

A core topic of economic discourse is the identification and promotion of robust strategies for industrial success. In the field of innovation studies, and at least since *Schumpeter*, private industry is largely perceived to be at the forefront of the process of technical change and that substantial resources are at stake when pushing forward the innovation frontier. This process is known to be complex, that is, characterized by strong non-linearities [42.1] and feedback between heterogeneous players [42.2]. Competitiveness is, thus, defined as a pattern of economic achievement having its roots in the micro-level decision-making of innovative firms in tradable sectors.

Different approaches are followed to understand, measure, and provide forward-looking strategic guidelines on the phenomenon of sustained industrial innovation. In the economics of technical change, much work deals with the impact of the level of R&D on economic performance (see, e. g., the contribution by [42.3]). In innovation studies, more broadly defined, it has been known for some years that large global innovative industrial corporations nurture portfolios of productive knowledge, which are weaved into ever more sophisticated products [42.4]. More recent discussions deal with the integration of services in manufacturing firm offerings, and new evidence points to increasingly complex business models that deliberately make services part of manufacturing activities [42.5–7]. A further line of enquiry is related to the pivotal role of softer dimensions of corporate positioning, namely marketing creativity, delivery characteristics, and design innovation that support and promote harder, functional, and technological propositions [42.8, 9].

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_42

All these unfolding combinations of elements certainly contribute to the economic advance of manufacturing enterprises, but it has been difficult to assess their specific relevance as effects are often described primarily qualitatively. Against this background, we take large sophisticated firms as bundles of capabilities [42.10, 11] and try to quantitatively capture some of these features with innovation indicators [42.12, 13]. We try to derive stylized facts for a few leading industrial/service firms advancing global offerings based on engineering excellence but also branding capabilities.

In Sect. 42.2, we present our basket of indicators to map and measure industrial profiles and trends. In Sect. 42.3, we present the data, the sample, and sources. In Sect. 42.4, the commonalities and particular trajectories characterizing the sample are identified. Finally, in Sect. 42.5, we conclude by arguing how our approach and findings matter to what is a topical and increasingly policy-relevant contemporary agenda.

42.2 The Approach to Studying the Interplay of Patents and Trademarks

In order to quantitatively assess the elements behind industrial success we collected a variety of data for a sample of European manufacturing firms. In our case, these are engaged in intermediate goods and capital equipment. The empirical material includes economic data on sales, profits, and R&D investment on the one hand, and the *output* innovation indicators on the other:

- Patent applications
- Trademark applications.

In our study, patent applications are taken to represent purposeful knowledge activities conducted with a view to launching new or improved technological-based artifacts. We are aware that in many cases, innovative activities are based on R&D, but also other origins of innovation are possible, e. g., the expertise of workers or ideas of clients. It is also clear that the proof of the ultimate value of patents is contingent on market outcomes and can hardly be inferred from the technologies' characteristics [42.14]. The pioneering work of *Grilliches* [42.15] and *Patel* and *Pavitt* [42.16] have made these and other advantages and limitations well known. The new developments of globalization and informatization undoubtedly cause stress in terms of analysis, but the indicator can be used with workable adjustments so as to yield useful and important insights [42.17–19].

Trademarks stand for customer-relevant symbols and corporate signs that make the offerings of a firm distinct, traceable, and accountable in the marketplace. These intangible assets emerge from deliberate efforts to build visibility, differentiation, and reputation. An advantage of this type of data is that it can reveal to what extent and in which directions services are being developed as part of the overall business model of a firm. From the early work by *Schmoch* [42.20] and *Mendonça* et al. [42.21] to the recent reviews by *Graham* et al. [42.22] and by *Schautschick* and *Greenhalgh* [42.23], we have learned that trademark data yield relevant information regarding innovation and differentiation.

As trademarks refer to symbols or signs and not technology, they are not submitted to an examination process. It can be assumed that a new mark is created in the context of a new product, and indeed it can be shown that trademarks are closely linked to innovation [42.20, 21]. On the one hand, compared to that of patents, the classification of trademarks is much less detailed, so that for technology-oriented analyses patents continue to be the better option. Trademarks, on the other hand, reflect the orientation on active marketing and new product prospection. A specific advantage of trademarks is that they cover not only tangible products, but also intangible ones. Innovative initiatives by knowledge-intensive services industries seem, indeed, well covered by trademark evidence [42.24, 25]. By the same token, in principle, the service-orientation in manufacturing could also be picked up. Unlike patents, it has been observed that trademarks have a comparative advantage in smaller firms and start ups [42.26, 27]. However, trademarks can track big corporations as well. Hence, this exercise attempts to provide an illustration of how the service activities of larger manufacturing R&D performers can be analyzed quantitatively through trademarks.

42.3 Empirical Basis of the Analysis

We collected economic data on sales, profits, and R&D for a sample of manufacturing enterprises from the *EU Industrial R&D Investment Scoreboard* of the top 1000 EU enterprises [42.28]. In this database, the key economic data for top R&D investors are compiled according to their absolute performance in R&D spending. Thus, by definition, all enterprises in our sample are R&D intensive. As the criterion is the absolute investment in R&D, the implication is also that the focus is on larger enterprises.

We collected data for the 10-year period 2005–2014 for four specific *medium-high-tech* sectors:

- *Automobiles and parts* (20 companies)
- *Health care equipment* (12)
- *Electronic equipment* (9)
- *Industrial engineering* (21).

During the collection, it proved to be decisive to identify enterprises that are included in the dataset for the complete 10-year period. That is to say, data for many enterprises are not available for all the years. In particular, and for reasons not documented in the source, in the health care equipment sector most enterprises appear only for a few years. Thus, in the end, the dataset was somewhat small but sufficient for rough statistical assessments. In the spirit of the Scoreboard source, one advantage of such a limited dataset is that it was possible to collect patent and trademark data for consolidated groups, i. e., in addition to the core enterprises, their affiliates, identified through their websites, were included. Another advantage of a managing a small sample was developing an awareness of the variety of individual strategies, which generally disappear in very large samples.

For these 62 enterprises, we collected the patent applications for the observation period. We analyzed patent applications and not granted patents, as the grants for the last years of our period are not yet available. Furthermore, grants do not necessarily represent higher values, as other elements are relevant for grants [42.14]. The analysis of applications proves to be more appropriate for reflecting innovation activities. In order to make the patent applications of enterprises from different nationalities comparable, transnational patent applications were analyzed that represented a sample of patent applications with higher economic value [42.29]. For the search, the database World Patents Index (WPI) in the version of the host STN (Science Technology Network) International was used, as it is based on patent families facilitating the search for transnational patents. For transnational patent applications, applications at the European Patent Office (EPO) or international applications at the World Intellectual Patent Organization (WIPO) are analyzed without double counting. Both types of applications are available in WPI (for more details, see the Chap. 37 in this Handbook).

For our sample, we also collected trademark applications. Trademark applications for the following national territories were compiled, ensuring that key continents were covered, in particular, North America (Canada, the United States), Latin America (Brazil, Mexico), Asia (China, Japan, Malaysia, South Korea), Europe (including Switzerland and Turkey). Since these kinds of industrial companies are export-oriented, cross-country trademarking information was compiled used community marks (for the 28 member states) and international patents (the variable number of countries being designated through the Madrid system, an official procedure for applying and managing registration worldwide). For the Nice class breakdown, we divided the period into two 5-year periods and isolated only a few types of marks (US, EU, and worldwide) in the sample of firms.

42.4 Assessment of Indicators

In this chapter, some key patterns of patents and trademarks in the sample analyzed are described to aid the understanding of the basic structures of the sample and to illustrate the long-term trends of the use of patents and trademarks.

42.4.1 Some Key Patterns of Patents in the Sample Analyzed

There are 88 020 patent applications overall in our database for the period 2004–2015. The aggregate of companies applied for a total of 7604 in the year 2004 and 9324 in 2015, an overall increase of 22.6% in this 10-year period. However, as depicted in Fig. 42.1, development over time was not smooth. There were years of vibrant dynamism in the beginning of the observation period and years of lethargic performance at the end, as well as moments of negative growth.

We can just note that the evolution of the technological achievement of the firms is connected to broader economic factors and external conditions. The short-term perturbation between 2007 and 2011 may be

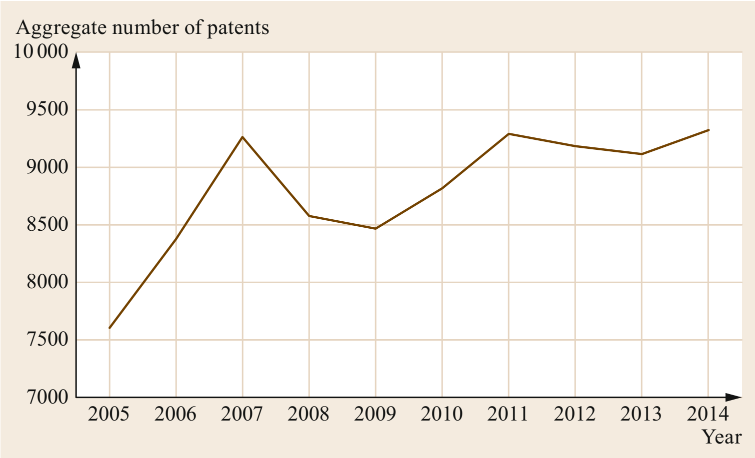


Fig. 42.1 Aggregate patent applications by the 62 firms in the sample. Source: WPI (STN), own computations

attributed to the global financial crises. From the patterns that emerge it is also possible to see that the slop of trend seems flatter after that, which hints at something of a long-term technological effect coming out of the economic crises. Only by 2014 did aggregate applications just surpass those of the peak year 2007. The ratio of patents to sales (m€) dropped from 22.6 in 2005–2009 to 19.5 in 2010–2014. It is as if economic austerity translated into technological austerity. This smaller emphasis on science and technology by companies weakened by adverse times may, of course, have rebound costs in terms of diminished future economic potential.

There is substantial variety in terms of industrial behavior (Fig. 42.2). For instance, the automobile and parts sector appears almost stagnant, whereas electronics is the most volatile. Moreover, health care equipment shows the most vibrant performance and, curiously, industrial engineering's strong growth was broken when the acute phase of the crisis was subsiding.

42.4.2 Some Key Patterns of Trademarks in the Sample Analyzed

There is a total stock of 12539 trademark applications in our dataset. This is much less than the number of patents, which shows that intermediate-input industries *do* trademark [42.20] but in much lesser volume compared with their patenting performance and compared with final consumer product industries [42.21]. There are, however, some interesting inter-industry variations (Table 42.1). For instance, the automobile sector appears to be quite patent *and* trademark driven, whereas electronics firms appear to emphasize both types of intellectual property rights less.

The evolution of trademark fillings was wobbly. As Fig. 42.3 shows, applications trended downwards in spite of a burst during 2011 and 2012. This burst was as large, but deflated rapidly. As Fig. 42.4 confirms, the trademark pattern contrasts with that of patents; patenting had a negative fluctuation from 2007 to 2011 but

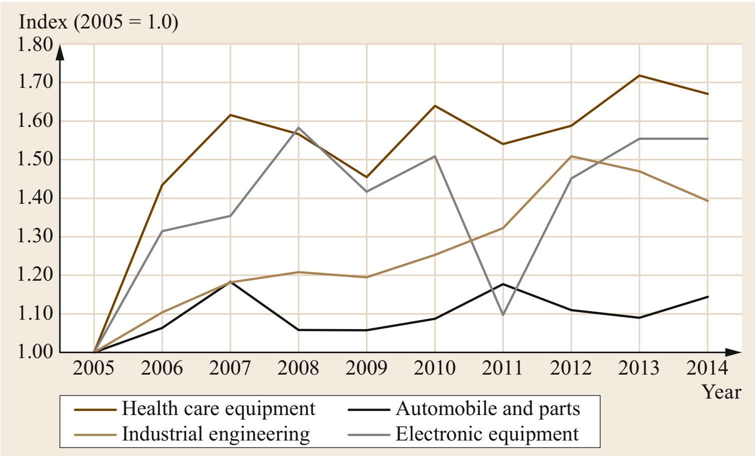


Fig. 42.2 Patenting per industry, index numbers (index base 2005)

Table 42.1 Distribution of total patents and trademarks by industry 2005–2014

	Patents (%)	Trademarks (%)
Automobile and parts	73	40
Health care equipment	9	29
Electronic equipment	3	8
Industrial engineering	15	23
	100	100

Table 42.2 Trademark applications and Nice class fillings by industry, percent change considering 2005–2009 and 2010–2014

	Applications (%)	Classes (%)
Automobile and parts	11.1	29.0
Health care equipment	−1.2	−34.1
Electronic equipment	0.4	−17.5
Industrial engineering	3.3	15.0

soon recovered, while trademarking jumped for a couple of years only to keep on sliding down, further distancing itself from the patenting trend. This observation is of interest, since research into the impact of the latest economic crisis is still unfolding [42.30, 31].

Certainly, these are aggregate figures; therefore insights are to be gained by breaking down per industry and type of trademark information. Some industries stagnated in terms of trademark applications, or even receded; only the auto sector pushed markedly up, 11.1% from 2005–2009 to 2010–2014 (Table 42.2).

In as far as trademarks are concerned, it is also revealing to consider class findings. In an analogue way to patents, the findings for trademarks are assorted into classes (the international Nice classification system) and typically to more than one class. These are selected by the applicants and carry a cost, which means this information is quantitatively and qualitatively rich. Table 42.2 shows that class requests tend to follow the signal of changes of trademark fillings (generally speaking, not always). For instance, the auto sector expanded applications but expanded the number of classes per trademarks even more, while the health technol-

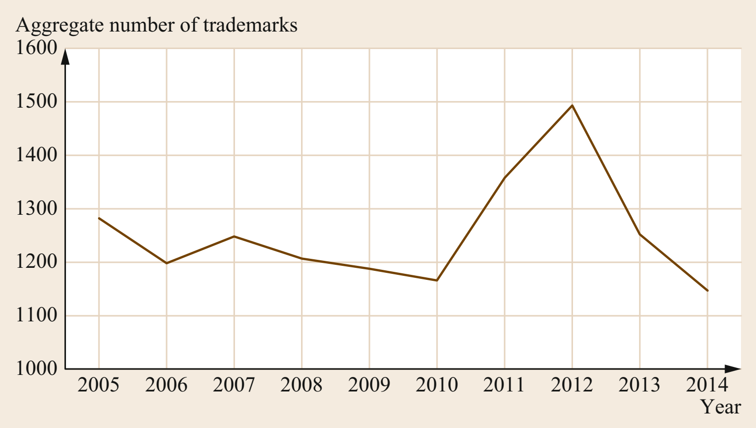


Fig. 42.3 Aggregate trademark applications by the 62 firms in the sample

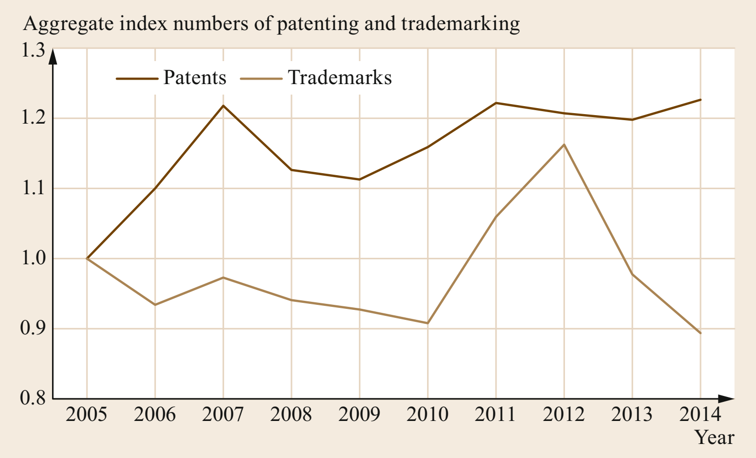


Fig. 42.4 Aggregate patenting and trademarking, index numbers

ogy sector diminished applications but even more so the number of classes per trademark.

Classes indeed yield another, finer level of analysis. It is possible to analyze, for instance, the dynamics of goods (classes 1–34) and services (classes 35–45). Pooling all the firms for all the years we find that 28.7% of all filled classes are service classes. As Table 42.3 shows, and as expected, there are different industry profiles and time dynamics. The health technology sector becomes the most service oriented, while the electronics business loses that status.

Next, we can re-classify class fillings according to their "sophistication" [42.32], especially to access if and to which extent manufacturing firms are active knowledge-based services [42.24]. In a way akin to the OECD (Organisation for Economic Co-operation and Development) typology, it is possible to re-organize Nice classes according their *technology* (in the case of goods) and *information* (in the case of services) *intensiveness*. This is done in Table 42.4.

From Table 42.4, it is clear that in our the sample, the firms' capabilities are mostly located in Nice classes that correspond to medium-high-technology (MHT) capabilities: 6214 trademark classes fillings of all the firms throughout the entire period. However, it should be borne in mind that these compose less than one-third of all the classes. This key stylized fact is important and is yet another instance underlining that *diversity* matters. It also reinforces previous observations that used patent as data and thus serve to portray these firms as *multi-knowledge corporations* (see [42.10, 11, 18, 33], which consider this a major advance of innovation studies in the last 20 years).

Table 42.3 Proportion (in percentages) of service trademarks in the overall portfolio of firms, by industry

	2005–2009	2010–2014
Automobile and parts	20.6	25.5
Health care equipment	35.6	36.2
Electronic equipment	43	36
Industrial engineering	28.0	27.9

It is notable that the second group of salient trademark classes are grouped into the *H-I* category. In fact, the single most trademarked individual class is Nice class 42 which refers to:

> Scientific and technological services and research and design relating thereto; industrial analysis and research services; design and development of computer hardware and software.

The other important services classes are classes 35 (business services) and 41 (education and recreation services). The nature of the complementarity between these types of service marketing resources with manufacturing capabilities is something worthy of further scrutiny (for a recent appraisal of the challenges of productive knowledge diversification, see [42.34]). This stylized fact suggests that the *profile* of knowledge configurations matter.

Finally, Table 42.4 also displays the dynamics of diversification as measured by corporate marketing resources. It suffices to note that the most sophisticated type of services are either the *most invested* into (the automobile sector, the engineering sector) or the least divested (the health care sector, the electronics sector). This trend is an important indication of the *direction* of diversification.

42.4.3 Patents, Marks, and Profitability

In this section, we will look at the link of patents and marks to profitability in order to understand which elements or combination of elements have an impact on profit. For this purpose, it is necessary to look at the specific characteristics of the dataset and run various analyses.

Key Techno-Economic Characteristics of the Sectors

The enterprises under analysis are part of four sectors characterized by high fluctuation and high competition. In the database of the EU scoreboard, the number of

Table 42.4 Scope of *sophistication* among the 62 firms, structure, and changes

	Pooled industry (absolute number)	Automobile and parts (change) (%)	Health care equipment (change) (%)	Electronic equipment (change) (%)	Industrial engineering (change) (%)
HT	5843	*28*	*−27*	*−8*	*17*
MHT	6214	*10*	*−23*	*−9*	*2*
MLT	1256	*48*	*−7*	*−38*	*35*
LT	1419	*47*	*−46*	*9*	*59*
H-I	3653	*75*	*−27*	*−26*	*45*
L-I	1850	*40*	*−56*	*−46*	*−13*

The *pooled industry* column displays all the data for all the companies for the period 2005–2014; percentages in *italics* are variations rates of class fillings between 2005–2009 and 2010–2014. *HT* high-tech, *MHT* medium-high-tech, *MLT* medium-low-tech, *LT* low-tech, *H-I* high-information-intensive, *L-I* low-information-intensive.

entries and exits within the four selected sectors is quite high. The variety within our dataset is still considerable, as the list of average, minimum, and maximum values for the different enterprise features clearly depicts (Table 42.5).

There are remarkable differences between the sectors along economic lines: The average profit rate (profits over sales) in health care is 13.0, in industrial engineering it is 7.8, or the average R&D rate is 8.2 in electronic equipment and 4.5 in industrial engineering, the average volume of sales in this period is 49.9 billion Euro in the automobile and parts sector, and 18.3 in electronic equipment. This variety shows that the conditions in the different sectors are quite specific, so that the values of the optimal economic performance for the different parameters may differ by sector.

Table 42.5 Features of the enterprise sample by sectors, 2005–2014

	Average	Min.	Max.
Total sample			
Sales	49.9	1.6	537.5
Profit rates	8.8	−31.8	27.3
R&D rate	5.4	1.1	14.9
Patents/sales	19.7	0.1	69.0
Product marks/sales	7.8	0.5	25.4
Service marks/sales	3.1	0.0	23.6
Automobile and parts			
Sales	117.4	4.7	537.5
Profit rates	5.8	−0.4	21.1
R&D rate	5.0	1.3	9.8
Patents/sales	17.1	0.3	42.0
Product marks/sales	5.0	0.5	23.7
Service marks/sales	1.5	0.1	5.1
Health care equipment			
Sales	31.1	2.5	143.0
Profit rates	13.0	2.4	20.9
R&D rate	6.2	1.8	11.4
Patents/sales	21.0	1.8	69.0
Product marks/sales	9.3	1.1	23.8
Service marks/sales	2.5	0.6	6.7
Electronic equipment			
Sales	18.3	3.2	65.7
Profit rates	8.3	2.0	20.4
R&D rate	8.2	1.1	14.9
Patents/sales	21.7	0.1	63.7
Product marks/sales	11.0	1.3	49.4
Service marks/sales	7.0	0.3	23.6
Industrial engineering			
Sales	35.6	1.6	89.9
Profit rates	7.8	−24.4	19.2
R&D rate	4.5	1.5	11.7
Patents/sales	20.6	4.7	60.6
Product marks/sales	8.2	0.8	25.4
Service marks/sales	3.2	0.5	18.3

Even within a sector, the parameters vary considerably, even for enterprises with similar products. A good example is that of producers of tyres, which are part of the automotive and parts sector. Tyre companies figure among the most vibrant and the most troubled of the automobile sector: Continental's sales grew by 136% between 2004–2009 and 2010–2014, while Pirelli practically stagnated at a rhythm of 4%.

Analysis of the Link of Profitability to Patents and Marks

For the analysis of the link of profitability—as a performance indicator—to patents and marks, the numbers of patents, product marks, and service marks were standardized by sales in order to compensate the different enterprise sizes. The computations were made for the 62 enterprises for the periods 2005–2009, 2010–2014, and the whole period 2005–2014. First we calculate specialization, e. g.,

$$\text{profit rate index} = \frac{\text{company profit rate}}{\text{average profit rate of the respective sector}}$$

to see whether the profit rate is above or below average. In order to eliminate the specific conditions in the four sectors, all data were normalized by the sector averages of the different parameters in the respective observation periods leading to a range of values between 0 and infinity, with 1 as a neutral value, i. e., values equivalent to the averages. As for this indicator, the values in the range of 0–1 are very concentrated; they were transformed to a symmetric scale, where the range is between −100 and +100 and the neutral value is 0.

$$\text{Trans profit rate index} = 100 * \text{tanhyp}(\ln(\text{profit rate index}))\,.$$

This procedure is equivalent to that used for specialization indices, see, e. g., [42.35, p. 7].

The following analyses were conducted:

1. Linear regressions between profit rates and patents/sales, product marks per sales, service marks per sales, total marks/sales, and share of service marks within total marks.
2. Lagged correlations between the parameters of 2005–2009 as to the profit rates of 2010–2014.
3. Limitation of the analyses on enterprises with profits above average in order to see whether these enterprises have specific features.
4. Analyses without outliers, i. e., without the upper and lower 5% of the cases, in order to check the robustness of the regression analyses.

As for linear regressions, the values were generally positive for 2005–2009, but on a very low level, and negative for 2010–2014. The lagged correlations did not lead to better results. The most positive regression results were achieved for enterprises with profits above average for the link between profit rates and product marks/sales and total marks/sales for the period 2005–2009 ($R^2 = 0.428$ and $R^2 = 0.423$, both with a significance level at 5%). This suggests that enterprises with high profit rates distinguish themselves from other enterprises by an explicit use of marks.

These analyses invite critical appraisal. The results may be explained by a few major reasons. First, it is reassuring that the correlation between R&D rate and patents/sales is positive with a coefficient of 0.5. Thus, there is a positive link, but other elements such as inventions based on sources other than R&D or different propensities of enterprises to patent play a role as well.

Second, the correlation between the variables and the profit rates are generally not linear, but rather quadratic. When doing the computations for the sample without outliers, the quadratic correlation for patents/sales exhibits an increase until the average level of 1 and a decrease for values above 1 (Fig. 42.5). This pattern tells a story of companies struggling if they have too few patents (perhaps not signaling that they are technological players) or too many (perhaps employing too many resources in strategic patenting, a behavior that could eventually be costly in terms of reputation). The implication is that there would be an optimal value for patents/sales. The relation seems to be the opposite for service marks over sales, meaning that some manufacturing companies can fare very well with a small number of very powerful service brands (say, ElringKlinger) or with a massive array of service brands (say, Nokian Tyres). The same outcome is found for the analysis with the non-parametric Lowess estimator, which approximates the sample locally. The LOWESS (locally weighted scatterplot smooting) estimator shows a roughly similar structure for product marks/sales, service marks/sales, and total marks/sales.

These correlations appear to be quite weak. This may be a consequence of the small sample of enterprises. With a larger sample, stronger correlations would probably have been found. Nevertheless, the small sample brings out the considerable variety in the behavior of firms, which would have been overlooked in a larger sample.

Finally, methodological caveats have to be taken into account. All enterprises of our sample may be considered as competitive, as they have survived for a longer period in competitive markets. Therefore, parameters other than profit rates may be more appropriate to reflect economic success, especially when there is evidence of curvilinear relationships [42.36].

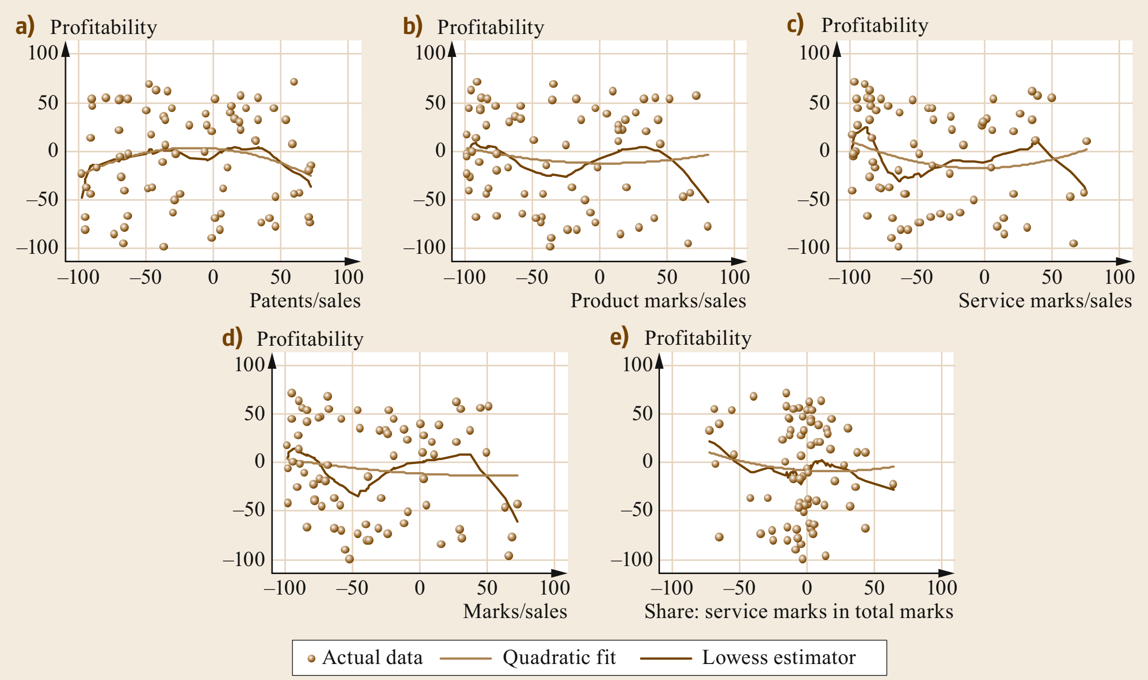


Fig. 42.5a–e Graphical representation of the relationship between profitability and various variables (outliers excluded). **(a)** Patents/sales; **(b)** product marks/sales; **(c)** service marks/sales; **(d)** marks/sales; **(e)** service marks in total marks

High absolute profits despite moderate profit rates due to a high level of sales can be an economic target as well. Another potential parameter can be the low variance of profit rates linked to profit rates above average. A further potential parameter may be a kind of *sharp ratio* or *reward-to-variability-ratio*, similar to this index in financial markets, which incorporates the profit rate in a certain period of time standardized by the absolute level of sales and the volatility.

Some Illustrative Examples

To illustrate the the connection between profit rates and the pattern of the explanatory features, some illustrative cases are presented. A striking observation for all parameters is that enterprises with the same values for, e. g., product marks achieve completely different profit rates. In detail, obviously, other strategic elements decide on the economic success. An interesting example is the producers of tyres whereof four are in the sector *automobiles and parts* (Table 42.6).

In this case, there are considerable differences in the R&D intensity, and these appear linked to that in patents/sales. For Continental, this can be explained by a broad product range, which encompasses tyres, automated driving, power trains, safety, etc. Michelin has a product range of automotive accessories, but the highest profit rate is achieved by Nokian Tyres with a low number of patents and a low R&D rate. The number of marks is higher than those of Michelin and Continental, but lower than that of Pirelli. The enterprise is specialized on winter tyres with high performance features. Compare the annual report 2016 of Nokian Tyres [42.37], where the orientation to tyres for extreme conditions is exposed as key strategy of the company. Thus, it aims at small, specialized niches with considerable success.

Another example comes from ElringKlinger, the features of which are shown in Table 42.7. ElringKlinger is a supplier of automotive parts in the area of motors and in-cabin elements. The profit rate is largely above average (in Table 42.7, the transferred indices in the range of −100 to +100 are used), but dropped a little bit between 2005–2009 and 2010–2014. The economic success is linked to a high R&D intensity and a high number of patents. The competitive strategy aims at high quality and innovation speed compared to competitors in Mexico, Hungary, or China, e. g., by developing tools for lightweight components (see the annual report of 2016 [42.38]). In 2005–2009, the level of product marks was above average and in 2010–2014, the level of service marks rose as well. Thus, in this specific case, the link of profitability to patents, product marks, and service marks is shown.

The enterprise Essilor is leading in the area ophthalmic optics and visual health; in addition to standard lenses, the company offers anti-fog, anti-reflective, polarized, or photochromic lenses (according to the annual report [42.39]). Again, the profit rates are largely above average. In this case, the activities in product and service marks have always been high (Table 42.8).

Table 42.6 Characteristic parameters for producers of tyres, 2005–2014

	Patents/sales	Product marks/sales	Service marks/sales	R&D intensity	Profitability	Sales
Nokian Tyres	0.3	6.8	5.1	1.3	21.1	11.5
Michelin	11.6	3.4	1.4	3.7	8.3	179.9
Pirelli	8.6	13.2	3.2	3.3	3.9	56.1
Continental	23.7	4.6	1.3	6.0	7.0	246.8

Table 42.7 Characteristic parameters of the ElringKlinger enterprise

Period	Patents/sales	Product marks/sales	Service marks/sales	R&D rate	Profitability
Basic values					
2005–2009	22.5	26.3	0.4	5.5	15.3
2010–2014	34.3	23.7	4.0	5.1	12.9
Index values					
2005–2009	27	89	−91	16	84
2010–2014	12	91	74	14	53

Table 42.8 Characteristic parameters of the Essilor enterprise

Period	Patents/sales	Product marks/sales	Service marks/sales	R&D rate	Profitability
Basic values					
2005–2009	24.1	14.8	11.1	4.7	17.6
2010–2014	15.4	6.7	15.6	3.2	18.8
Index values					
2005–2009	−5	5	26	−17	38
2010–2014	−17	9	12	−54	27

Table 42.9 Characteristic parameters of the Kone enterprise

Period	Patents/sales	Product marks/sales	Service marks/sales	R&D rate	Profitability
Basic values					
2005–2009	12.3	4.8	2.5	1.3	15.2
2010–2014	13.5	1.7	4.8	1.4	15.3
Index values					
2005–2009	−47	−34	−88	−79	68
2010–2014	−17	9	12	−54	27

Table 42.10 Characteristic parameters of the TomTom enterprise

Period	Patents/sales	Product marks/sales	Service marks/sales	R&D rate	Profitability
Basic values					
2005–2009	41.0	11.8	7.0	5.3	3.2
2010–2014	28.6	8.5	22.7	15.6	−2.1
Index values					
2005–2009	36	−5	−49	−29	−74
2010–2014	36	−18	6	59	−100

However, high profits are not always linked to high patent or R&D activities, trademarks may make the difference. The company Kone, specialist in elevators, has a profile as shown in Table 42.9. The profit rate is very high, but in 2005–2009 all parameters were below average. The R&D and patent levels remained moderate in 2010–2014, however, the activities in marks rose remarkably. The strategy aims the orientation at the needs of the customers, which is also linked to new technologies, but at a moderate level (see the annual review of 2016 [42.40]).

In the case of the enterprise TomTom, specialized in navigation systems, the profit rate was moderate in 2005–2009 and even became negative in 2010–2014 (Table 42.10). Especially since 2012, it has raised its R&D intensity substantially, obviously to cope with the requirements in a very competitive market. However, it may be that the high investment in R&D is the reason for the actual losses, notwithstanding being a major asset for the next years. Thus, the company is performing R&D in the context of autonomous driving, a key area of the next years (see the annual report of 2016 [42.41]). In any case, the number of patents and service marks increased in parallel to the up-take of R&D.

The different examples show that the link of profits to patent, trademark, or R&D activities is not so simple and has to be embedded into the context of specific markets and enterprise histories [42.42]. Overall, the different parameters depend on corporate and industrial trajectories, but the analysis reveals that this industry is indeed exhibiting a [42.43, p. 270]:

> changing mix of innovation from traditional *hard* science-based research toward a greater emphasis on *softer* competencies in design and trademarks.

42.5 Conclusions

Contemporary companies from intermediate manufacturing industries nurture diversified knowledge portfolios and are increasingly offering integrated *product-service* solutions. Capturing these trends and understanding the patterns calls for a new wave of innovation studies and techno-metric work that combines the indicators patents and trademarks, while distinguishing between goods and service marks, as well as gradations of sophistication within mark categories [42.44, 45].

The optimal configurations of capabilities that are relevant for the economic success of manufacturing enterprises are not straightforward, as they are contingent of the market context and individual strategies. Patents have a positive impact on profits until a certain maximum level, which depends on the specific situation in the respective sector. In any case, a minimum level of patents is necessary to achieve economic success.

Trademarks also emerge as relevant for the economic success of upstream industries, and the relevance of these assets increased between the two observation periods 2005–2009 and 2010–2014. The most crucial point is that our findings point to the fact that the use of service marks augmented considerably and, in particular, especially those in knowledge-base classes. The knowledge base of large and export-oriented manufacturing firms is distributed along technological resources but also covers particular marketing capabilities.

The core thesis of this chapter, that this, the emphasis that a growing service orientation of manufacturing firms matters for performance, must be further tested and buttressed. More work can and should be invested in linking service sophistication to particular industrial profiles. One promising methodological step could be the building of composite empirical constructs, combining degrees of tangibility and sophistication. Marks and softer kinds of innovation are increasingly important, and further work should be channeled to this area, especially to consider more downstream sectors as well [42.45–47].

With respect to strategy, bundling and unbundling technological and marketing resources is likely to remain challenging. The definition of economic success by profit rates has obvious shortcomings and further research is necessary to achieve satisfying concepts. Plus, the conditions of learning resilience that enable companies to withstand economic crises certainly deserve further study. Finally, this multifaceted agenda may be of benefit to policy-makers, as modern productive organizations are evolving combinations of resources, which are unlikely to respond linearly to linear or unidimensional strands of policy.

References

42.1 S.J. Kline, N. Rosenberg: An overview of innovation. In: *The Positive Sum Strategy: Harnessing Technology for Economic Growth*, ed. by R. Landau, N. Rosenberg (National Academy Press, Washington 1986) pp. 275–305

42.2 C. Freeman: *Systems of Innovation: Selected Essays in Evolutionary Economics* (Edward Elgar, Cheltenham 2008)

42.3 A.B. Jaffe: Technological opportunity and spillovers of R&D: Evidence from firms' patents, profits, and market value, Am. Econ. Rev. **76**(5), 984–1001 (1986)

42.4 K. Pavitt: Innovation process. In: *The Oxford Handbook of Innovation*, ed. by J. Fagerberg, D. Mowery, R.R. Nelson (Oxford Univ. Press, Oxford 2004) pp. 86–114

42.5 T.S. Baines, H.W. Lightfoot, O. Benedettini, J.M. Kay: The servitization of manufacturing: A review of literature and reflection on future challenges, J. Manuf. Technol. Manag. **20**(5), 547–567 (2009)

42.6 W. Reim, V. Parida, D. Örtqvist: Product-service Systems (PSS) business models and tactics—A systematic literature review, J. Clean. Prod. **97**, 61–75 (2015)

42.7 F. Adrodegari, N. Saccani: Business models for the service transformation of industrial firms, Serv. Ind. J. **37**(1), 57–83 (2017)

42.8 M.L. Swink, M. Song: Effects of marketing-manufacturing integration on new product development time and competitive advantage, J. Oper. Manag. **25**(1), 203–217 (2007)

42.9 P. Stoneman: *Soft Innovation: Economics, Product Aesthetics, and the Creative Industries* (Oxford Univ. Press, Oxford 2010)

42.10 O. Granstrand, P. Patel, K. Pavitt: Multi-technology corporations: Why they have 'distributed' rather than 'distinctive core' competencies, Calif. Manag. Rev. **39**(4), 8–25 (1997)

42.11 G.N. von Tunzelmann: Localized technological search and multi-technology companies, Econ. Innov. New Technol. **6**(2/3), 231–256 (1998)

42.12 C. Freeman (Ed.): *Output Measurement in Science and Technology. Essays in Honor of Yvan Fabian* (North Holland, Amsterdam 1987)

42.13 H.F. Moed, W. Glänzel, U. Schmoch (Eds.): *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems* (Springer, Dordrecht 2004)

42.14 R. Frietsch, U. Schmoch, B. van Looy, J.P. Walsh, R. Devroede, M. du Plessis, T. Jung, Y. Meng, P. Neuhäusler, B. Peeters, T. Schubert: *The Value and Indicator Function of Patents* (EFI, Berlin 2010)

42.15 Z. Grilliches: Patent statistics as economic indicators: A survey. In: *R&D and Productivity: The Econometric Evidence*, ed. by Z. Griliches (Univ. of Chicago Press, Chicago 1998) pp. 287–343, paper originally published in 1990

42.16 P. Patel, K. Pavitt: The continuing, widespread (and neglected) importance of improvements in mechanical technologies, Res. Policy **23**(5), 533–545 (1994)

42.17 H. Grupp, U. Schmoch: Patent statistics in the age of globalisation: New legal procedures, new analytical methods, new economic interpretation, Res. Policy **28**(4), 377–396 (1999)

42.18 S. Mendonça: The revolution within: ICT and the shifting knowledge base of the world's largest companies, Econ. Innov. New Technol. **15**(8), 777–799 (2006)

42.19 P. Neuhäusler, R. Frietsch: Patent families as macro level patent value indicators: Applying weights to account for market differences, Scientometrics **96**(1), 27–49 (2013)

42.20 U. Schmoch: Service marks as novel innovation indicator, Res. Eval. **12**(2), 149–156 (2003)

42.21 S. Mendonça, T.S. Pereira, M.M. Godinho: Trademarks as an indicator of innovation and industrial change, Res. Policy **33**(9), 1385–1404 (2004)

42.22 S.J.H. Graham, G. Hancock, A.C. Marco, A.F. Myers: The USPTO trademark case files dataset: Descriptions, lessons, and insights, J. Econ. Manag. Strategy **22**(4), 669–705 (2013)

42.23 P. Schautschick, C. Greenhalgh: Empirical studies of trade marks: The existing economic literature, Econ. Innov. New Technol. **25**(4), 358–390 (2016)
42.24 U. Schmoch, S. Gauch: Service marks as indicators for innovation in knowledge-based services, Res. Eval. **18**(4), 323–335 (2009)
42.25 S. Mendonça: Trademarks as a telecommunications indicator for industrial analysis and policy. In: *Telecommunication Economics*, Lecture Notes in Computer Science, Vol. 7216, ed. by A.M. Hadjiantonis, B. Stiller (Springer, Berlin, Heidelberg 2012)
42.26 M. Flikkema, A.P. de Man, C. Castaldi: Are trademark counts a valid indicator of innovation? Results of an in-depth study of new Benelux trademarks filed by SMEs, Ind. Innov. **21**(4), 310–331 (2014)
42.27 G. de Vries, E. Pennings, J.H. Block, C. Fisch: Trademark or patent? The effects of market concentration, customer type and venture capital financing on start-ups' initial IP applications, Ind. Innov. **24**(4), 325–345 (2017)
42.28 European Commission: The 2016 EU Industrial R&D Investment Scoreboard, http://iri.jrc.ec.europa.eu/scoreboard16.html (2017)
42.29 R. Frietsch, U. Schmoch: Transnational patents and international markets, Scientometrics **85**, 185–200 (2010)
42.30 R. Cabral, M. Alves, R.P. Mamede, S. Mendonça, J. Seixas, V. Soromenho-Marques: *Heading South: Rethinking the Eurozone – The Ulysses Study* (Heinrich-Böll-Stiftung European Union, Brussels 2018)
42.31 J. Caraça, B.-Å. Lundvall, S. Mendonça: The changing role of science in the innovation process: From Queen to Cinderella?, Technol. Forecast. Soc. Change **76**(6), 861–867 (2009)
42.32 S. Mendonça (Ed.): *Remark: Estudo Sobre o Contributo das Marcas Para o Crescimento Económico e Para a Competitividade Internacional* (INPI, Lisbon 2011)
42.33 B.R. Martin: Twenty challenges for innovation studies, Sci. Public Policy **43**(3), 432–450 (2016)
42.34 M. Nasiriyar, L. Nesta, L. Dibiaggio: The moderating role of the complementary nature of technological resources in the diversification–performance relationship, Ind. Corp. Change **23**(5), 1357–1380 (2014)
42.35 U. Schmoch, C. Michels, P. Neuhäusler, N. Schulze: *Performance and Structures of the German Science System 2011, Germany in International Comparison, China's Profile, Behaviour of German Authors, Comparison of Web of Science and SCOPUS*, Studien zum deutschen Innovationssystem EFI, Nr. 9-2012 (EFI, Berlin 2012)
42.36 T. Pohlmann, P. Neuhäusler, K. Blind: Standard essential patents to boost financial returns, R&D Manag. **46**(S2), 612–630 (2015)
42.37 Nokian Tyres: Annual report 2016, http://annualreport2016.nokiantyres.com/ (2017)
42.38 ElringKlinger: Annual report 2016, http://ar2016.elringklinger.de/downloads/en/elringklinger-ar-2016.pdf (2017)
42.39 Essilor: We should all care for our eyes, 2016–2017 annual report, Charenton-le-Pont – France, https://www.essilor.com/annualreport2016/pdf/ESSILOR_RA2016-2017_EN.pdf (2017)
42.40 Kone: Annual review 2016, https://www.kone.com/en/Images/KONE_Annual_Review_2016_tcm17-37391.pdf (2017)
42.41 TomTom: Annual report 2016, https://corporate.tomtom.com/static-files/765e69fb-27ef-4ae2-8ec0-a415c3a12d20 (2017)
42.42 M.M. Godinho, V. Ferreira: Analyzing the evidence of an IPR take-off in China and India, Res. Policy **41**(3), 499–511 (2012)
42.43 V.G. Scalera, D. Mukherjee, A. Perri, R. Mudambi: The role of innovation in inventory turnover performance, Multinatl. Bus. Rev. **22**(3), 270–293 (2014)
42.44 J. Hagedoorn, M. Cloodt: Measuring innovative performance: Is there an advantage in using multiple indicators?, Res. Policy **32**(8), 1365–1379 (2003)
42.45 S. Mendonça: National adaptive advantages: Soft innovation and marketing capabilities in periods of crisis and change. In: *Structural Change, Competitiveness and Industrial Policy: Painful Lessons from the European Periphery*, ed. by A. Teixeira, E. Silva, R. Mamede (Routledge, London 2014) pp. 133–150
42.46 M. Napoletano, L. Nesta: The multifaceted evolution of industries: Sectoral and firm-level evidence, Ind. Corp. Change **25**(5), 809–812 (2006)
42.47 M. Spring, L. Araujo: Service, services and products: Rethinking operations strategy, Int. J. Oper. Prod. Manag. **29**(5), 444–467 (2009)

Sandro Mendonça

ISCTE University Institute of Lisbon
Lisboa, Portugal
sfm@iscte.pt

Sandro Mendonça is Professor at the Department of Economics, ISCTE. His research is focused on innovation, industrial policy and economic history.
He is faculty of the University of Sussex, Invited Professor of ISEG, and also supervises the doctoral programs of the Southern Medical University and the University of Electronic Science and Technology of China. Since early 2018 he has accepted to serve as a member of the board of the Portuguese national authority for communications, ANACOM.

Ulrich Schmoch

Fraunhofer-Institut für System- und Innovationsforschung ISI
Karlsruhe, Germany
ulrich.schmoch@isi.fraunhofer.de

Professor Ulrich Schmoch holds a Doctoral degree in Social and Political Sciences (University of Hannover, Germany), a Diploma of Mechanical Engineering (University of Hannover) with modules at the University of Grenoble, France, and a State Doctorate in Sociology of Science and Technology (Karlsruhe Institute of Technology KIT, Germany). He has been working in different functions at the Fraunhofer Institute for Systems and Innovation Research (ISI), Karlsruhe, Germany, since 1986. Ulrich Schmoch has carried out and coordinated about 200 projects in the areas of innovation indicators, patent analysis, bibliometrics, systems of innovation, technology transfer, science management, innovation in threshold countries, innovations in the service sector, science-based technologies, and organization and mission of universities in an international comparison.

Peter Neuhäusler

Competence Center Policy – Industry – Innovation
Fraunhofer Institute for Systems and Innovation Research ISI
Karlsruhe, Germany
peter.neuhaeusler@isi.fraunhofer.de

Peter Neuhäusler is a Project Manager at the Competence Center Policy - Industry - Innovation at the Fraunhofer Institute for Systems and Innovation Research (Karlsruhe). He studied Social Sciences, majoring in Micro- and Macro-Sociology, Applied Social Research, and Quantitative Methodology at the University of Mannheim and graduated with a diploma in 2008. Peter Neuhäusler was awarded his Doctoral degree in Innovation Economics at the Technical University of Berlin in 2012.

43. Post Catch-up Trajectories: Publishing and Patenting Activities of China and Korea

Chan-Yuan Wong, Hon-Ngen Fung

This chapter seeks to explore the sequential cyclical growth of science, technology, and science-based technology for two economies—China and South Korea—in the course of transitioning to the postcatching-up phase. Both China and South Korea intend to capitalize on scientific and technological knowledge in order to transition to the postcatching-up phase of development. This chapter highlights the production trajectories of science and technology towards the postcatching-up phase in terms of:

1. Scientific publications
2. Granted patents
3. Copatenting pattern
4. Forward citations
5. Science-based patents.

China and South Korea have been active in terms of scientific publication and patenting activities. In regard to patenting, both economies have shown the capability to produce patents and are able to converge the growth of patents with that of publications. This chapter highlights a generic cyclical growth path for science, technology, and science-based technology in the course of transitioning to an advanced knowledge-based economy. It is nonetheless important to explore if there are different paths pursued by other emerging economies.

43.1 Background

The *BRICS* countries along with the *next 11* economies are recognized as nations with promising outlooks for investment and economic growth. These emerging economies are acknowledged to be both dynamic and quickly catching up with the industrialized economies with South Korea and China, in particular, close to achieving advanced industrialization. The scientific and technological catching-up process of these two economies has been a subject of great interest to many development economists and innovation scholars [43.1–3].

South Korea stands out as one of the most successful economies, especially in terms of scientific and technological capability. It achieved this through capitalizing on its resources to support conglomerate firms, which invested in and expanded productive industries during the early catching-up phase. Upgrading and exporting are the essential activities that South Korean firms must carry out in order to receive support from the state. The expansion of many South Korean conglomerates led to massive employment, human resource and skill development, income distribution, and—as a result—socioeconomic development [43.4]. There are many studies explicating how firms from different industries evolved and what catch-up strategies were adopted—first to learn process capabilities, and then to develop sophisticated technologies via highly organized R&D and brand development [43.5–9]. The size of the firms and collective upgrading efforts (well-coordinated activities organized by the state) allowed South Korean firms to be able to compete with other advanced economies in certain capital-intensive industries such as automotive, telecommunications, iron and steel, and construction.

South Korea acquired productive routines for manufacturing industries and benefited from the export mar-

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_43

ket. Many South Korean firms leveraged these gains and utilized the growing local technological market as a test bed to explore niches [43.10, 11]. South Korea saw concerted efforts to develop their scientific and technological knowledge stock for advanced science-based industrial development [43.8]. It is reported that there is high participation of productive firms in academically oriented activities. South Korea has witnessed growth of a dynamic and extensive network in its innovation system which would prove instrumental for knowledge-based industrial development [43.12, 13].

Conversely, China has experienced significant industrial development since the economic reforms in 1978 [43.14], with particularly noticeable development in the coastal regions. The industrial policy of China was instrumental during the early catching-up phase, attracting foreign direct investment (FDI) from multinational corporations that capitalized on labor-intensive manufacturing industries. China then pursued systematic reforms in its production and institutional structures in order to develop indigenous technologies. It instituted a policy experimentation routine to execute different approaches to economic and industrial catching-up strategies. The routine seeks to manage complex challenges in appropriating lessons from the successes of one sector or region for application in others [43.15–17]. There was a wide range of support from the state to create an institutional routine for R&D and to build linkages between productive firms, public research institutions (PRIs), and universities. Indeed, China has witnessed tremendous state-targeted R&D programs in recent years [43.3] to acquire an institutional *path-creating* routine, a routine of adopting emerging technologies for industrial upgrading [43.18], which allows for the emergence of niches. China has also witnessed significant growth in scientific publications and technological patenting since the early 2000s and mid-2000s, respectively [43.8, 19], with a significant percentage of R&D investment over GDP. While South Korea manifested relatively advanced innovation systems that were instrumental for knowledge-based industrial development, it is expected that China will rely on the momentum from growing knowledge-based industries, as its commitment to science and technology is quite evident.

Both South Korea and China intend to capitalize on scientific and technological knowledge to transition to the postcatching-up phase of development. Many studies highlighted their attempts to mobilize resources and create intellectual capital for advanced technology development [43.9, 19–23].

In the case of South Korea, there has been consistent support from the government to advance scientific and technological capability. Commitment to building research infrastructure started early (since the 1970s), and favors scientists and engineers from universities and research institutes who publish their findings or patent their inventions. It is noted that joint research activities between industrial engineers and academics are not uncommon in South Korea and such activities have been mandated by government-sponsored research programs since the 1980s [43.21, pp. 642–643]. The programs target research areas which would empower industrial technologies, and have indeed supported several leading firms in the development of frontier technological products. This in turn stimulated more joint research activities in the postcatching-up period (since 2000) and the demand for basic research to develop science-based technologies in universities and public research institutions [43.22].

In addition, there has been commitment from the state to empower small- and medium-sized enterprises (SMEs) with scientific and technological knowledge. There are incentives and support for SMEs to participate in joint activities with research and development (R&D) consortiums. In addition, the state established a public procurement system for emerging technologies developed by SMEs or start-ups. Through patenting analysis, *Doh* and *Kim* [43.1] found that the supports for SMEs established by the government are effective in building strong research relationships in Korea's networked economy.

In the case of China, it is noted that funding for R&D has been growing since 1997 and it is expected that this commitment to R&D will allow China to match the European Commission target of 3% of GDP expenditure in R&D. The increase in funding is seen as a factor that will *push* the increase in publications. China established a reward system under the long-term plan for the Development of Science and Technology [43.24] to incentivize scientists and academics who can perform academic publishing and patenting activities. In addition, *Zhou* and *Leydesdorff* [43.25] noted that there has been an increase in Chinese returnees who studied abroad, contributing to the stock of scientific knowledge (publications). They also observed that China has been playing an important role in publishing in nanorelated sciences, and this can be seen as a building block for nanotechnology development. It is also noted that policies shifted dramatically since 2003 (particularly in technology)—from horizontal to direct government interventions—in order to achieve the desired industrial structure. Since 2006, many megaprojects targeting emerging industries such as biotechnology and new materials were launched, and massive government stimuli were disbursed in 2009 to ensure such projects were implemented even during the global financial crisis. The projects indeed shaped scientific and technological activities via a policy process developed in a systemic manner that connected the vision of national leaders with stakeholders such as local governments, firms, and interest groups [43.3].

It is noted that both economies witnessed a substantial rise in scientific publications and patenting activities. This chapter highlights the production trajectories of science and technology towards the postcatching-up phase in terms of:

1. Scientific publications
2. Granted patents
3. Copatenting patterns
4. Forward citations
5. Science-based patents.

We wish to explore if there is a growth pattern for science, technology, and science-based technology in the course of a transition towards the postcatching-up phase. Thus, the perspective of convergence between scientific and technological activities to achieve science-based technology development is taken into consideration in the analysis. The development will be assessed from aspects of both quantity and quality (impact in terms of citations) of publishing and patenting trends.

Our systematic review of science, technology, and science-based technology production in South Korea and China will provide insights and a possible understanding of cyclical development in the transition towards the postcatching-up phase. The term *cycle* denotes an illustrative period stretching from the entry to the decline of a particular production [43.18, 26]. In what follows, a cycle refers to a specific production (science or technology) that would cover project entry, take-off, maturity, and ultimately the achievement of carrying capacity. The production of Japanese scientific papers illustrated in Fig. 43.1 can be seen as a typical example of a science production cycle for an economy. Japan witnessed publication entry in the early 1990s and take-off in 1998, ultimately producing at its carrying capacity since the early 2000s.

Meyer [43.27] and *Wong* and *Goh* [43.28] maintained further that a process of systemic change in a science (or technology) production may occur where a new institutional setting outgrows, replaces, or coevolves with the old one. A new cycle may emerge if the *push*-oriented production process (producing science for specific applications or industrial technology, ultimately witnessing dual flow in advancing science and technology) is linked to production that is experiencing *pull* by market demand forces [43.28]. The concomitance between science and technology is projected in *Bernadas* and *Albuquerque* [43.29].

We seek to provide a model in the broad sense, as a stylized framework of science and technology development processes emphasizing an approach to advanced development. Our assessment of publishing and patenting activities in the economies of China and South Korea will provide a useful guide for policymakers in other emerging economies who aspire to creating similar trajectories for their own advanced phase of development.

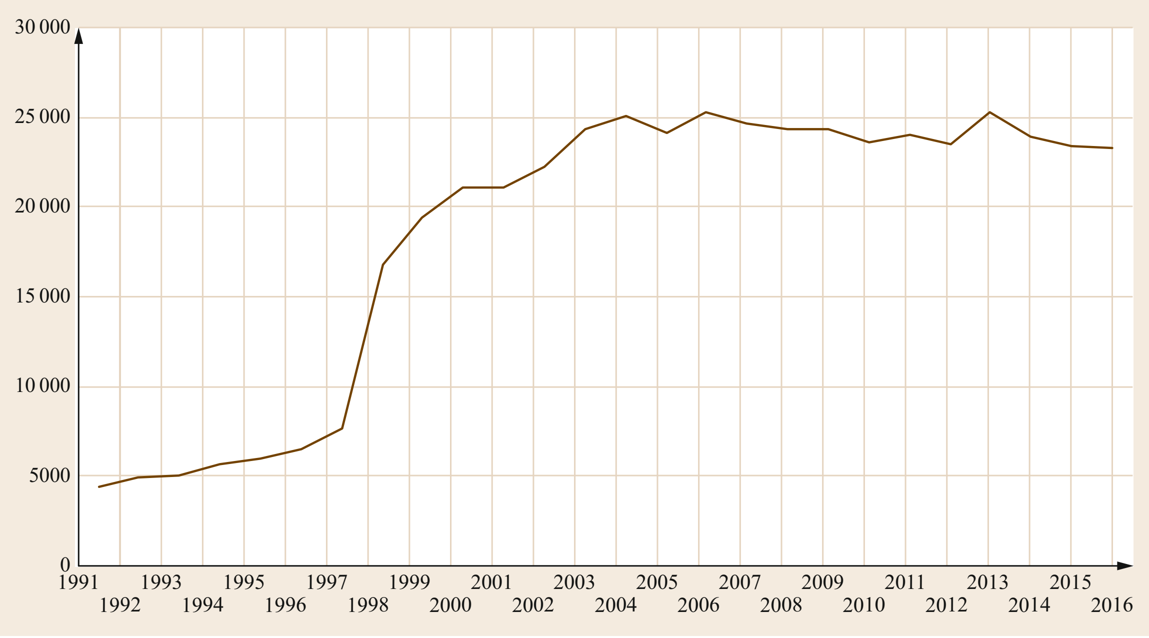


Fig. 43.1 Scientific publications of Japan. Data is extracted from the Web of Science (WoS; SCIE and SSCI) accessed on 23 July, 2017

43.2 Conceptual Framework and Data

43.2.1 Narrative Framework

Science-push and *market-pull* has been a subject of discussion among innovation scholars and it is useful as a narrative model in many policy and management frameworks. The implications of *science-push* with sequential *pull* informs policy-makers about how to mobilize resources for science (scientific knowledge) in an economy. The commitment would in turn create a spillover effect in the economy. The model provides the aspiration that the resultant scientific knowledge would be translated into technological knowledge (by the agents of change in the innovation system). The translation process would lead to the emergence of new industries and niches [43.28, 30].

The translation process is explained in the literature as a necessary link in any commitment to bridge scientific knowledge with technological knowledge. Many economies established intermediaries [43.31, 32] to ensure that the knowledge generated from universities and public research institutions is codified and applicable—that is, capable of diffusing services and has market demand [43.33, 34]. Areas of modern technology—such as biotechnology and nanotechnology—in large part emerged from the collective efforts to link scientific knowledge with industrial applications. The implication of *market-pull* basically informs policy-makers about the importance of research consortium networks in an economy that supports/allows capable entrepreneurs to translate scientific knowledge into useful applications—as they search for and define niches and create (socio-)economic value [43.32]. The ultimately increasing returns would fuel the interaction between science, technology, and economic growth. *Bernadas* and *Albuquerque* [43.29] suggested a comprehensive model that postulates the stages of development of science, technology, and economic growth (Fig. 43.2). Such a virtuous cycle of development is evident in many advanced economies. In addition, there are many emerging economies that show signs of attaining a highly functional structure that would fuel science, technology, and growth.

Schmoch [43.30] depicts the typical order of development for science-based technologies. He observed that there are basically two cycles of development before the emergence of market demand (the third cycle) and realization (diffusion) of a particular technology (product). While the two preceding cycles are projected via publishing and patenting trends, the third cycle is projected via the sales trend of the particular technology. It can be argued that the positive interaction between science and technology is triggered after a breakthrough in output of scientific articles [43.28]. *Javenpaa* et al. [43.35] extended the study of *Schmoch* [43.30] in illustrating cycle sequences for different technologies. There are technologies developed via a sequence of concomitant cycles as observed by *Schmoch* [43.30], and there are other technologies that diverge from the

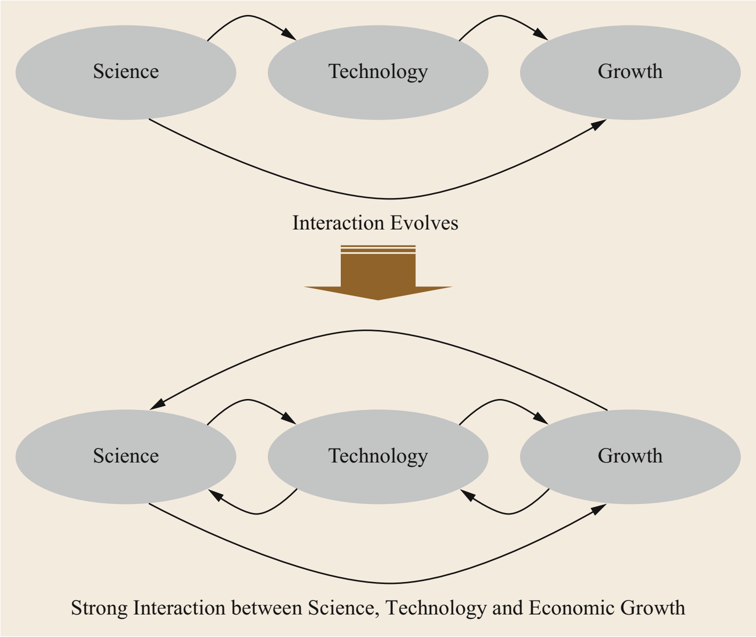


Fig. 43.2 The evolution of interaction between science, technology, and growth. Adapted from [43.29] with permission from Elsevier

typical sequence. This is particularly evident when technologies are explored via a nonscience route (e. g., tinkering).

43.2.2 Data

This chapter seeks to explore the typical type of scientific and technological development attained by emerging economies. The publishing and patenting performance of China and South Korea are used as indicators to reveal the science and technology trajectories acquired by economies that are in transition towards the postcatching-up phase. We are interested in highlighting whether the selected economies attained a concomitant cycle that would fuel science and technology activities for knowledge-economic development.

The process of scientific exploration and technology development has been relied upon as a means to unearth new discoveries and potential solutions to industrial and social issues. This study derives its content from scientific articles (publications) and patents. Scientific publications indicate the production of credible scientific knowledge and includes articles, reviews, and conference papers. The rigorous peer-review process which accompanies each scientific publication would also indicate credibility and transparency on the part of the knowledge producer. Some studies used scientific publications as an indicator to reflect the size and quality of the university education system. It enables one to connect the production of scientific publications with the provision of highly skilled graduates. Conversely, the production of patents indicates the accumulation of novel, inventive applications for industry; this sheds light on the innovative capacity of the subject. A subset of patents covered in this chapter is characterized by their citation to publications (as other references), which indicates the scientific base that exists at the foundation of the patent. We conjecture that there will be an increase in science-based patents for both economies.

The historical series of publications listed in Scopus and utility patents granted by the US Patent and Trademark Office (USPTO) have been extracted for this study. This chapter utilized Scival as a means to extract data from Scopus regarding China and South Korea. Scival is a part of the Elsevier research management suite and acts as an analytics platform for Scopus data after 1996, used in this context to capture the number of publications and forward citations of articles emerging from South Korea and China between 1996 and 2015 [43.36]. While we find the Scopus database instrumental for us to assess the historical series of publications, it is noted that the citation rates of the selected economies are in contrast to that of Clarivate Analytics' Web of Science, which may be due to the contrasting data structures during the early building-up stage. The chapter also categorized the data set based on subject areas and subcategories as defined by the Scopus Classification which covers 27 major subject areas and 313 subcategories. The CWTS (Center for Science and Technology Studies) Classification of individual journals is also utilized to identify the orientation and target audience of the various journals that authors from South Korea and China most commonly publish in [43.37].

The patent data is derived from the USPTO through Patsnap, a patent search engine that covers various patent offices worldwide. For the purposes of this chapter, patents were extracted for assignees originating from South Korea or China between the years 1996 and 2015. In this case, the assignees were classified based on keywords in their assignee name to determine whether they are categorized as university or industry. While we managed to extract valuable data to develop our analysis, we observed a general decline phenomenon for patent citations of many economies since the early 2000s. This may be attributable to the high workload of the USPTO's search engine and thus it is unable to provide us with accurate citation reports.

This chapter also sought to classify the patents extracted in order to assess the level of research and development intensity (RDI). The first step of this process is the concordance of the patent subclasses based on the concordance for IPC8 to NACE Rev. 2 [43.38]. This is followed by a series of steps to translate the patent subclasses to ISIC Rev 3.1 [43.39]. This process allows for the patents to be classified based on the classification released by OECD (Organization for Economic Co-operation and Development), which groups different product classes based on R&D expenditure divided by value added, over R&D expenditure divided by production [43.40]. This process is an extension of prior attempts to develop a coarse concordance between patent classes and product classes [43.41, 42].

The evolution of networks established for joint patenting activities are studied in detail. We review the literature on how the selected economies match their scientific resources with technological activities, and assess the effectiveness of their matching efforts by projecting the convergence process between scientific publications and patenting activities. The chapter utilized VantagePoint—a text mining software developed by Search Technology Inc.—to conduct the joint patenting mapping exercise with the focus on the top 30 patent assignees. The mapping is done by measuring the similarity index between patent records by assignees (node) in the various fields, which determines the thickness of the lines (edges) between the nodes.

43.3 Findings and Discussion

43.3.1 Science

Figure 43.3 shows the scientific publication performance of the two economies. China—a large emerging economy with rich resource endowment for science—is performing on a much larger scale in scientific publication. China witnessed its take-off point in 2003, growing from 24.6 papers per million population in 1996, to 125.7 in 2006, and achieving 269.6 in 2015. It is believed that there was a general push for science—as in many emerging countries—and the production has grown exponentially since.

Although the growth of China eclipsed that of the other relatively advanced economy, South Korea, the latter's production of scientific publications is nonetheless noteworthy for this study, as its take-off point appears to be similar to that of China. South Korea—a highly capable economy in performing high tech innovation since the 1980s (when many South Korean business conglomerates believed that the semiconductor industry was the basis of every other industry and stepped up their investment to upgrade capabilities for semiconductor technological innovations [43.43])—attained a much steeper growth curve before its take-off point as compared to China. The production of scientific publications by South Korea grew at a much quicker rate after the take-off point, from 192.4 papers per million population in 1996, to 580.2 in 2006, and reaching 1253.9 in 2015. The momentous growth in publications of the two economies may be due to the arrival of the respective take-off phase in their publishing cycles. This may be attributable to the commitment of the economies to incentivize scientists and researchers from universities to publish their research findings. Conversely, the universities are keen to recruit productive scientists to produce scientific papers. They also enforced the *publish or perish* performance criteria to evaluate the performance of faculty. This led to a significant increase in scientific publications from the emerging countries such as China and South Korea. Many publishing houses for scientific results have since then expanded their production to accommodate the demand.

In addition, the funding for academic research activities has been favorable and substantial particularly for those who are involved in the targeted fields [43.3, 20]. It seems that their resources (both in terms of hu-

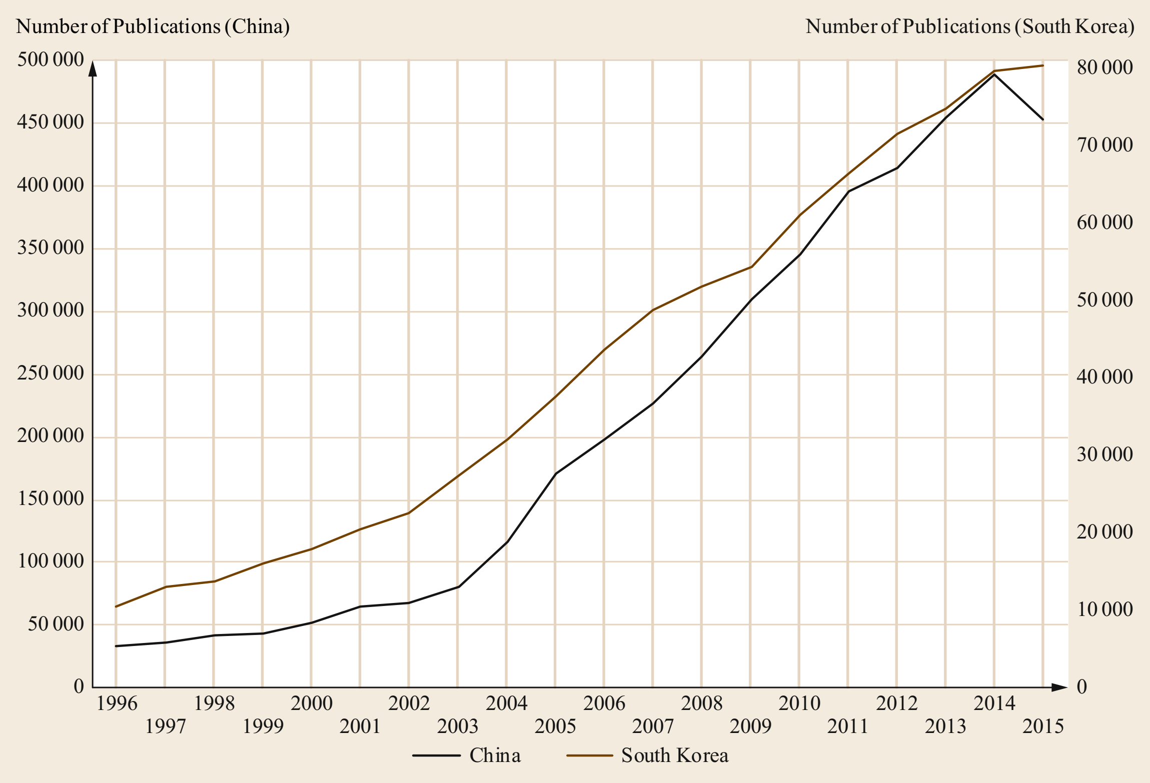


Fig. 43.3 Scientific publications of China and South Korea

man capital and finance) for science are far from being exhausted, and thus, it is unlikely the economies will achieve their carrying capacity soon.

The publications of the two economies prior to the take-off year saw a tremendous impact as measured by citations per article (Fig. 43.4). For the case of South

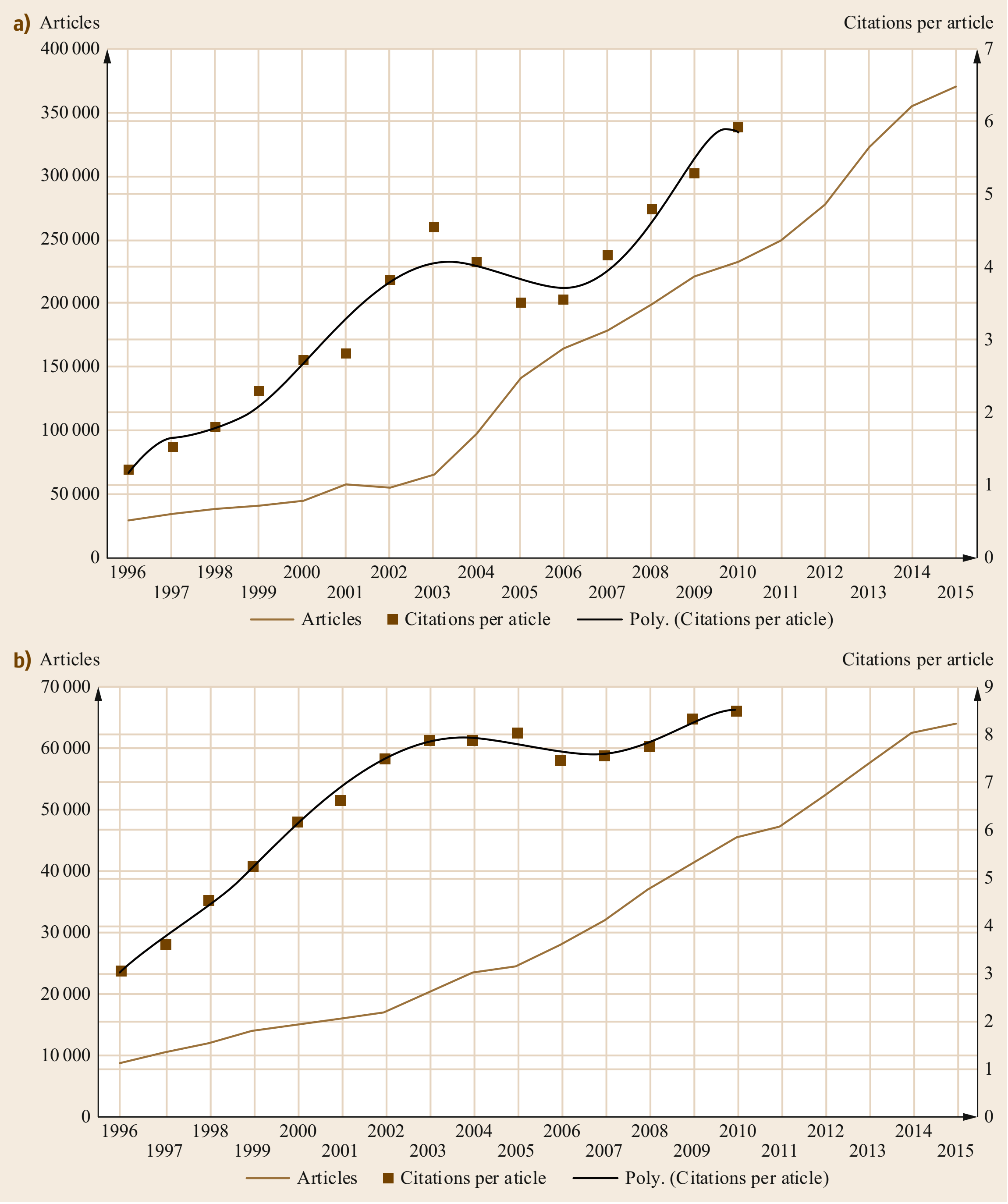


Fig. 43.4a,b Publications (in terms of articles in journals) and citations per article of China **(a)** and South Korea **(b)**. Yearly citation counts are adjusted to a 5 year citation window. This is due to the assumption that a citation cycle of a publication would end by the close of a 5 year period. *Bornmann* and *Leydesdorff* [43.44] used citation counts of 3–5 years to study the research landscape. The citations per article trend is plotted from 1996–2010

Korea, citations since then reached their first cyclical peak in 2003–2005. The cycle of high-impact scientific articles witnessed a saturation point and received relatively lower citations per article in the following years. The significant citations count by the early 2000s can be attributed to the *push* factor for scientific publications. Many articles published during the take-off year would tend to cite articles published in the previous five years, hence the citations trend is a normal phenomenon as it reflects the systematic conduct of the research process. The *quantifiable* aspect of impact for each article has seen a recovery since 2008 and this may due to the transition towards the *pull* phenomenon, which implies the significant application of science. For the case of China, its citations per article trendline showed a noticeable hump-shaped curve. Again, this reflects a cycle of citations in its publication trend. The upward citations rate started to resume after the take-off point for publications that emerged in 2003. It is noted that the trends for citations that are based on Web of Science data are noticeably different from those of this study. We conjecture that the contrasting trends can be attributed to the contrasting data structures of the two databases.

In terms of subject areas, chemical engineering- or chemical science-related areas tend to dominate the landscape of top publications of the two economies (Table 43.1). Chemical engineering with subcategories of *Colloid and Surface* and *Chemistry Catalysis* emerged as the top areas in terms of citations per publication. China seems to also have given research priority to other scientific fields such as pharmaceutical-related, material-related, and energy- and environment-related areas. In addition, China is also productive in *archaeology*, a subarea of *arts and humanities*. While South Korea also performed in energy- and environment-related and material-related areas, it also targeted multidisciplinary-related (which saw tremendous growth and a high number of authors involved), medicine (clinical genetics in particular), and nursing as top research priorities.

Figure 43.5 provides an overview of the share of joint publications between universities and industries for science. This reflects the joint commitment of two different entities in the innovation system to produce science for potential applications. Conversely, China witnessed an increase in its share of joint publications (all types), from less than 1% in 2003 (the take-off period) to about 1.3–1.6% afterwards. This shows a commitment to bridging science and industrial applications. Conversely, South Korea performed at a relatively higher level of joint publications within

Table 43.1 Top publications by subject areas, 2011–2015 (ranked by citations per publication)

Subject area	Subcategory	Publications	Publications (growth %)	Citations	Authors	Citations per publication
China						
Chemical engineering	Colloid and surface chemistry	6126	34.2	167 313	20 155	27.3
Chemical engineering	Catalysis	24 274	86.5	533 947	60 481	22
Pharmacology, toxicology and pharmaceutics	Pharmacology, toxicology and pharmaceutics (miscellaneous)	252	1085.7	3974	1287	15.8
Materials science	Biomaterials	14 410	86.1	218 223	45 355	15.1
Arts and humanities	Archaeology (arts and humanities)	385	259.3	5757	946	15
Chemistry	Electrochemistry	16 288	73.5	239 383	43 118	14.7
Environmental science	Ecological modeling	1645	63.3	23 283	4886	14.2
Chemical engineering	Bioengineering	23 921	47.2	310 557	69 276	13
Energy	Renewable energy, sustainability and the environment	26 630	173.4	336 266	63 420	12.6
Chemistry	General chemistry	134 359	64.8	1 620 300	255 259	12.1
South Korea						
Chemical engineering	Colloid and surface chemistry	1230	−19.7	33 551	3187	27.3
Chemical engineering	Catalysis	3662	31.3	80 249	8389	21.9
Multidisciplinary	Multidisciplinary	4079	490.8	61 663	10 812	15.1
Environmental science	Environmental chemistry	3569	48	49 568	7378	13.9
Chemistry	Electrochemistry	3111	−4.7	42 782	7193	13.8
Medicine	Genetics (clinical)	785	29.2	10 628	2963	13.5
Materials science	Biomaterials	3971	43.1	52 015	10 075	13.1
Chemical engineering	Filtration and separation	530	49.4	6735	1217	12.7
Energy	Renewable energy, sustainability and the environment	5320	77.9	66 180	10 825	12.4
Nursing	Advanced and specialized nursing	321	22.6	3982	1020	12.4

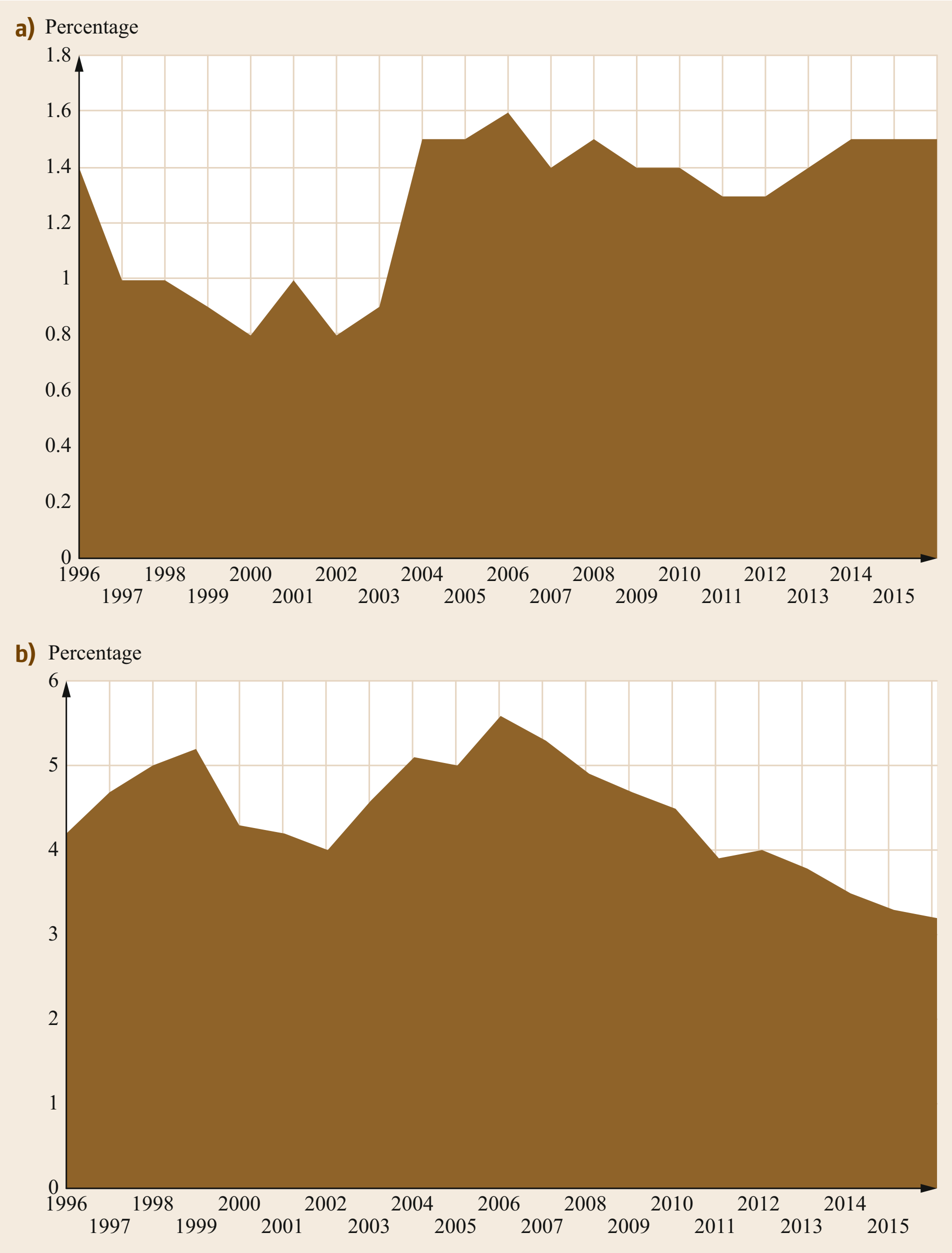


Fig. 43.5a,b Copublications between academia and industry (%). **(a)** China; **(b)** South Korea

the 3–6% range—for comparison, in 2015 Germany attained a share of 3.5%, France 3.2%, Denmark 5.8%, and Italy 1.9%. South Korea's commitment to joint research activities began early (in the 1980s), attaining a highly functional innovation system by 1996 when its share of joint publications was higher than that of China. It is noted that collaboration between scientists from universities and public research institutes and industrial engineers is not uncommon in the context of South Korea. It is also noteworthy that South Korea's share has been declining since 2007. *Kwon* et al. [43.45] maintained that the cross-authorship within entities in Korea has been steadily eroding as universities and industries seek to internationalize their collaboration structures. We on the other hand, conjecture that this phenomenon can be attributed to recent efforts of many universities in advanced economies to venture into emerging science and technology. The knowledge may be too niche for industrial players to identify its immediate applications, hence many universities are commercializing it via university start-up mechanisms.

Table 43.2 shows the top journals by publication counts for the two economies. The journal titles are listed in the Scopus database, and many of the titles are also listed in the WoS. We sought to match the WoS's journal titles with Leiden University's CWTS field orientations of different journals to identity the targeted readership of the journals [43.37]. There are journals oriented towards advancing scientific knowledge among the peers of a particular field, and there are journals established to inform medical or industrial practitioners about the potential applications of new sciences. Table 43.2 lists the orientation of journal titles that are listed in the WoS.

China produced 225 965 publications in the top 20 journals of China. We observed that 69% of the publications are also listed in the WoS. Many of these publications are intended for both internal circulation (within scientific communities of China) and dissemination abroad. There is a substantial number of publications that are meant for circulation among peers in a particular field. China produced 85 720 publications that are meant for university-oriented science. Publications for application-oriented science account to about 45% of the total WoS-listed publications in top journals of China. The science *push* commitment of China seems to have an impact on university-oriented science publications. As China continues to bridge science for new applications, it is expected that the orientation will shift towards application-oriented science in the near future.

Conversely, South Korea's publications manifested a slightly different development pattern. It produced 68 267 publications in its top 20 journals, of which 91% are listed in the WoS. From the total WoS listed publications, 89% are oriented towards medical or industrial applications. This implies that South Korea has achieved a concomitant stage that empowers a virtuous cycle of science and technology development. Most of the scientific publications that South Korea produces have industrial applications, while it still seeks to retain some commitment to advancing basic science (university-oriented). This can be attributed to the early commitment of the South Korean government in bridging science (produced by the academics) and technology for industry. The effort to advance application-oriented science would ultimately advance science-based technologies of South Korea.

43.3.2 Technology

Many studies have reported on the significant efforts made by China and South Korea in terms of indigenous building technology [43.4, 6, 15, 16], advancing capabilities for high-tech industries, securing IP protections for their inventions, and branding their innovations for markets abroad [43.9, 19]. They have both been active in patenting, with their patenting trend showing significant growth in the past few years. China witnessed its take-off point for patenting in 2009, about 6 years after its take-off point for publishing activities—going from just 0.02 patents per million population in 1996, to 0.3 in 2006, and reaching 5.3 in 2015. Conversely, South Korea witnessed its take-off point for patenting in 2005 and achieved growth comparable to that of its publications before convergence in 2014 (Fig. 43.6). It witnessed a capacity of 31.1 patents per million population in 1996, expanded to 119.3 in 2006, and reached 360.5 in 2015. The phenomenal growth of patenting in the two economies appears to echo that of publications. It is noted that the two economies are keen to routinize patenting activities to develop identified technologies that would spawn niches in the global production value chain.

Both economies—but China in particular—achieved significant reductions in the gap between publications and patents. The ratio between publications and patents for China in 1996 was 1385 : 1, reduced to 744 : 1 in 2000, further reduced to 168 : 1 in 2010, and reaching 62 : 1 in 2015. This reflects the national commitment toward correlating development of science with technology, which is particularly crucial for the catching-up economies. The importance is in order for catching-up economies to be able to translate scientific results into technology that empowers socioeconomic activities [43.29]. In this respect, South Korea achieved a low gap in the 1990s, attaining a ratio of 7 : 1 in 1996 and ultimately reducing it further to just 4 : 1 in 2015.

For patenting impact, South Korea seems to have witnessed a (full) cycle of patent citations—from a rapid increase of citation counts in the 1990s, to plateauing in the early 2000s, and subsequently declining since then (Fig. 43.7). For the patent citations, we utilize 5 year patent citation windows. While the decline of citations may imply a waning of the patenting impact of Korea, it may also be attributable to the high workload of USPTO's search engine making it thus unable to provide accurate citation figures for the examination report.

Nonetheless, we also note the decline in citations per patent over time, which we conjecture is due to ventures into long-cycle types of emerging technologies (e. g., science-based technologies such as medicine) [43.46, pp. 45–69] and [43.47, p. 51]. The ventures have yet to achieve significant impacts (in terms of citations) for these economies. We also ob-

Table 43.2 Top 20 journals by number of publications, CWTS field orientations, 1996–2015

Journals	Count	Listed in WoS/Scopus	Orientation (university/industrial/medical)
China			
PLOS One	23 156	WoS/Scopus	Medical applications oriented science
Acta Physica Sinica	18 321	WoS/Scopus	University-oriented science
Acta Crystallographica Section E Structure Reports Online	15 949	WoS/Scopus	University-oriented science
RSC Advances	12 176	WoS/Scopus	University-oriented science
Chinese Physics Letters	12 139	WoS/Scopus	University-oriented science
Chinese Journal of Clinical Rehabilitation	12 001	Scopus	—
Proceedings of the Chinese Society of Electrical Engineering	10 819	Scopus	—
Applied Physics Letters	10 813	WoS/Scopus	Industrial application oriented science
Journal of System Simulation	10 340	Scopus	—
Journal of Alloys and Compounds	9880	WoS/Scopus	Industrial application oriented science
Transactions of the Chinese Society of Agricultural Engineering	9805	Scopus	—
Chinese Science Bulletin	9386	WoS/Scopus	University-oriented science
Materials Letters	9253	WoS/Scopus	University-oriented science
Journal of Functional Materials	9209	Scopus	—
Acta Optica Sinica	9188	Scopus	—
Spectroscopy and Spectral Analysis	9138	WoS/Scopus	Industrial application oriented science
Journal of Applied Polymer Science	8830	WoS/Scopus	Industrial application oriented science
Rare Metal Materials and Engineering	8618	WoS/Scopus	Industrial application oriented science
Chinese Medical Journal	8496	WoS/Scopus	University-oriented science
Acta Electronica Sinica	8439	Scopus	—
Total	225 956		
Total publications listed in WoS	156 155		
Total university oriented science publications	85 720		
Total industrial/medical applications oriented science	70 435		
South Korea			
Journal of the Korean Physical Society	9090	WoS/Scopus	Industrial application oriented science
Bulletin of the Korean Chemical Society	8186	WoS/Scopus	Industrial application oriented science
Applied Physics Letters	5309	WoS/Scopus	Industrial application oriented science
Journal of Korean Medical Science	3652	WoS/Scopus	Industrial application oriented science
Korean Journal of Dermatology	3518	Scopus	—
Biochemical and Biophysical Research Communications	3323	WoS/Scopus	Industrial application oriented science
PLOS One	3312	WoS/Scopus	Medical applications oriented science
Journal of Microbiology and Biotechnology	3124	WoS/Scopus	Industrial application oriented science
Journal of Applied Physics	2712	WoS/Scopus	Industrial application oriented science
Korean Journal of Chemical Engineering	2665	WoS/Scopus	Industrial application oriented science
Physical Review B Condensed Matter and Materials Physics	2590	WoS/Scopus	University-oriented science
Electronics Letters	2584	WoS/Scopus	Industrial application oriented science
Archives of Pharmacal Research	2556	WoS/Scopus	Industrial application oriented science
Journal of Mechanical Science and Technology	2514	WoS/Scopus	Industrial application oriented science
Journal of Nanoscience and Nanotechnology	2336	WoS/Scopus	Industrial application oriented science
Journal of the Korean Society of Food Science and Nutrition	2305	Scopus	—
Yonsei Medical Journal	2189	WoS/Scopus	University-oriented science
Lecture Notes in Computer Science including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics	2148	WoS/Scopus	Industrial application oriented science
Japanese Journal of Applied Physics: Part 1 Regular Papers and Short Notes and Review Papers	2078	WoS/Scopus	Industrial application oriented science
Molecules and Cells	2076	WoS/Scopus	University-oriented science
Total	68 267		
Total publication listed in WoS	62 444		
Total university oriented science publications	6855		
Total industrial/medical applications oriented science	55 589		

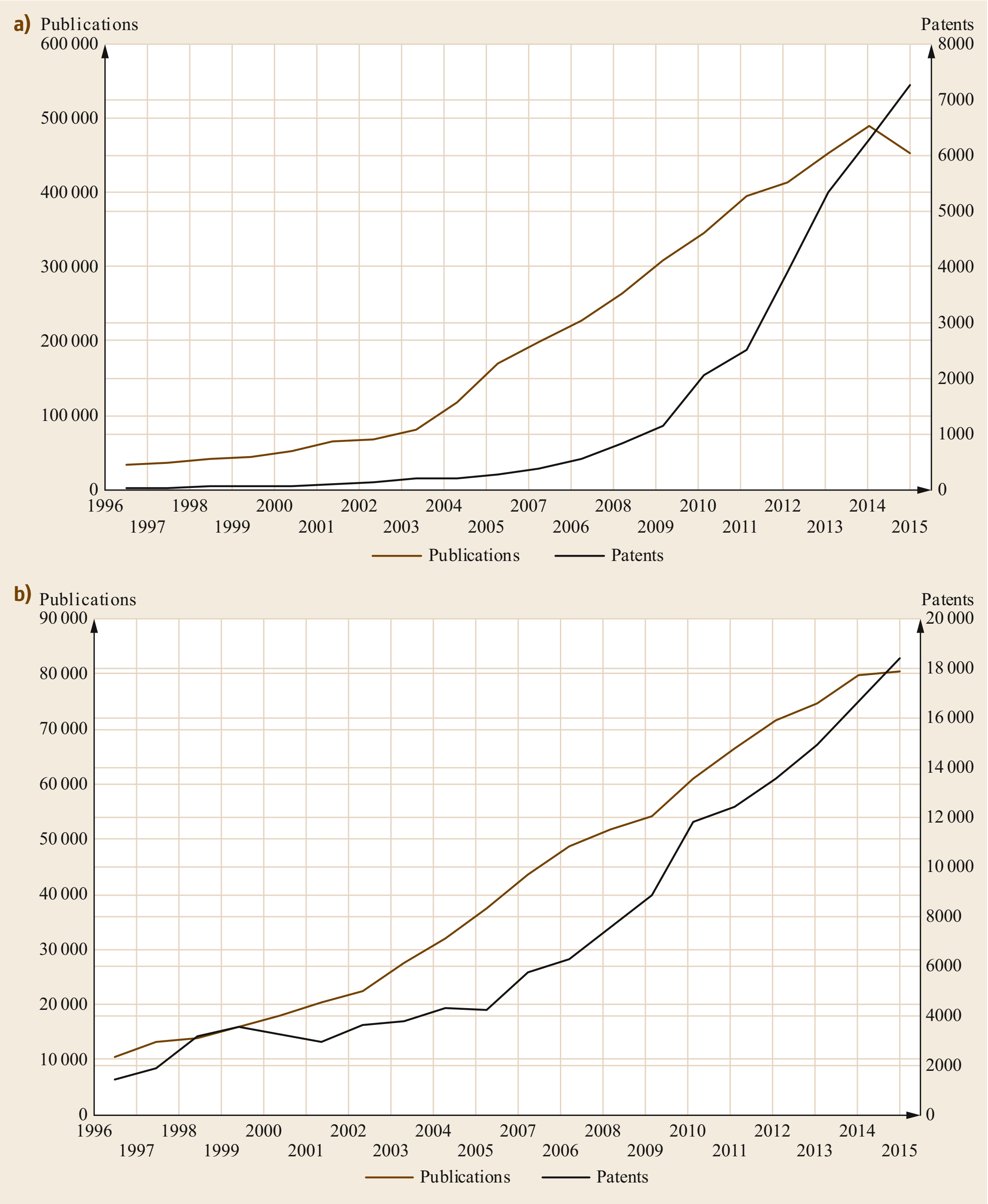


Fig. 43.6a,b Publications and patents of China (**a**) and South Korea (**b**)

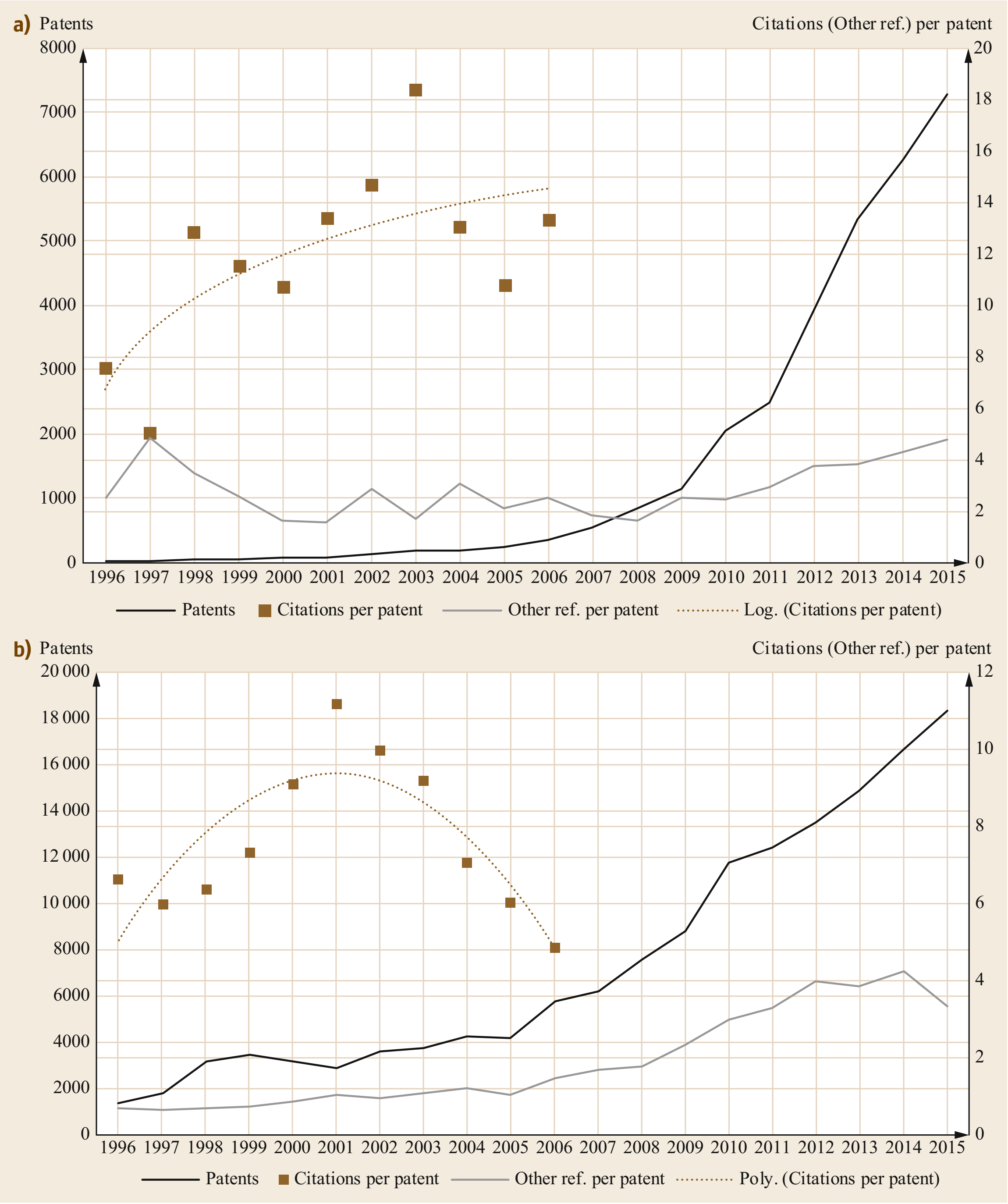


Fig. 43.7a,b Patents, citations per patent, and other references per patent of China **(a)** and South Korea **(b)**. Note: Yearly citation counts are adjusted to 5 years citation window. The citations per patent trend is plotted from 1996–2006. The increasing delay between priority date and search report impeded our attempts to track recent figures for citations per patent

Table 43.3 Technological level of patenting

	China					South Korea				
	High tech	Medium high tech	Medium low tech	Low tech	Outside of manufacturing	High tech	Medium high tech	Medium low tech	Low tech	Outside of manufacturing
1996	30 (34%)	29 (33%)	16 (18%)	14 (16%)	0	1170 (36%)	1154 (36%)	698 (22%)	206 (6%)	11 (0.3%)
2005	180 (32%)	186 (33%)	129 (23%)	73 (13%)	0	3518 (37%)	3307 (35%)	2090 (22%)	515 (6%)	10 (0.1%)
2015	5822 (38%)	5424 (36%)	2664 (18%)	1203 (8%)	0	14 904 (38%)	14 820 (38%)	7268 (19%)	2038 (5%)	0

served a significant increase of other references (i. e., not patents) backward citations per patent since the take-off points of the respective economies. Other references backward citations reflect possible input of scientific knowledge in crafting an invention. The patenting performance of South Korea seems to have caused a *pull* for scientific knowledge for new inventions or creation of new applications or niches. Conversely, China witnessed a consistent increase in patenting impact.

Table 43.3 shows the translation results of concordance between patent classifications and the level of technology. We observe that the selected economies have been keen to patent their inventions that are categorized as high tech and medium high tech. The share of patents in the two levels of technology has increased, while the share of low tech has fallen. This is evident particularly for the case of China. The *pull* for scientific knowledge in these two economies can be attributed to the pursuit of high tech patenting. The *pull* factor and significant efforts of these economies to bridge science and technology have indeed led to the increase of co-owned patenting between those from the universities and those from industry. Table 43.4 shows a significant increase in terms of co-owned counts over the decades in the selected economies. China's copatenting counts increased from virtually none in 1996, to 277 in 2015. Meanwhile, South Korea witnessed an increase from 1 to 335 in the same period of time.

Figure 43.8 shows the copatenting networks of China and Korea, which are mapped based on data of the top 30 assignees of each economy, extracted from

Table 43.4 Co-owned patenting between university and industry of China and South Korea

	China		South Korea	
	U	U-I	U	U-I
1996	4	0	1	1
2005	23	13	35	17
2015	644	277	659	335

the respective patent documents. We observed that there are different entities featured in the network. China's patenting landscape is dominated by many private firms and there are a number of network links for copatenting activities. Public research institutions such as the Chinese Academy of Sciences and China Academy of Telecommunications Technology are featured in the landscape. Universities such as Peking University and Tsinghua University are also featured in the landscape and their patents are coventured with those from the industries.

By contrast, South Korea manifested a much wider network for copatenting activities and the landscape is dominated by private firms, with Samsung Electronics emerging as the key player in the map. Many universities and PRIs are visibly featured and their influences in technology are evident. Korea Advanced Institute for Science and Technology (KAIST) emerged as one of the prominent influential players in the patenting landscape. It has links connecting private firms and PRIs for copatenting activities. This echoes our earlier observation on the concomitant stage for science and technology attained by South Korea for knowledge-based economic development.

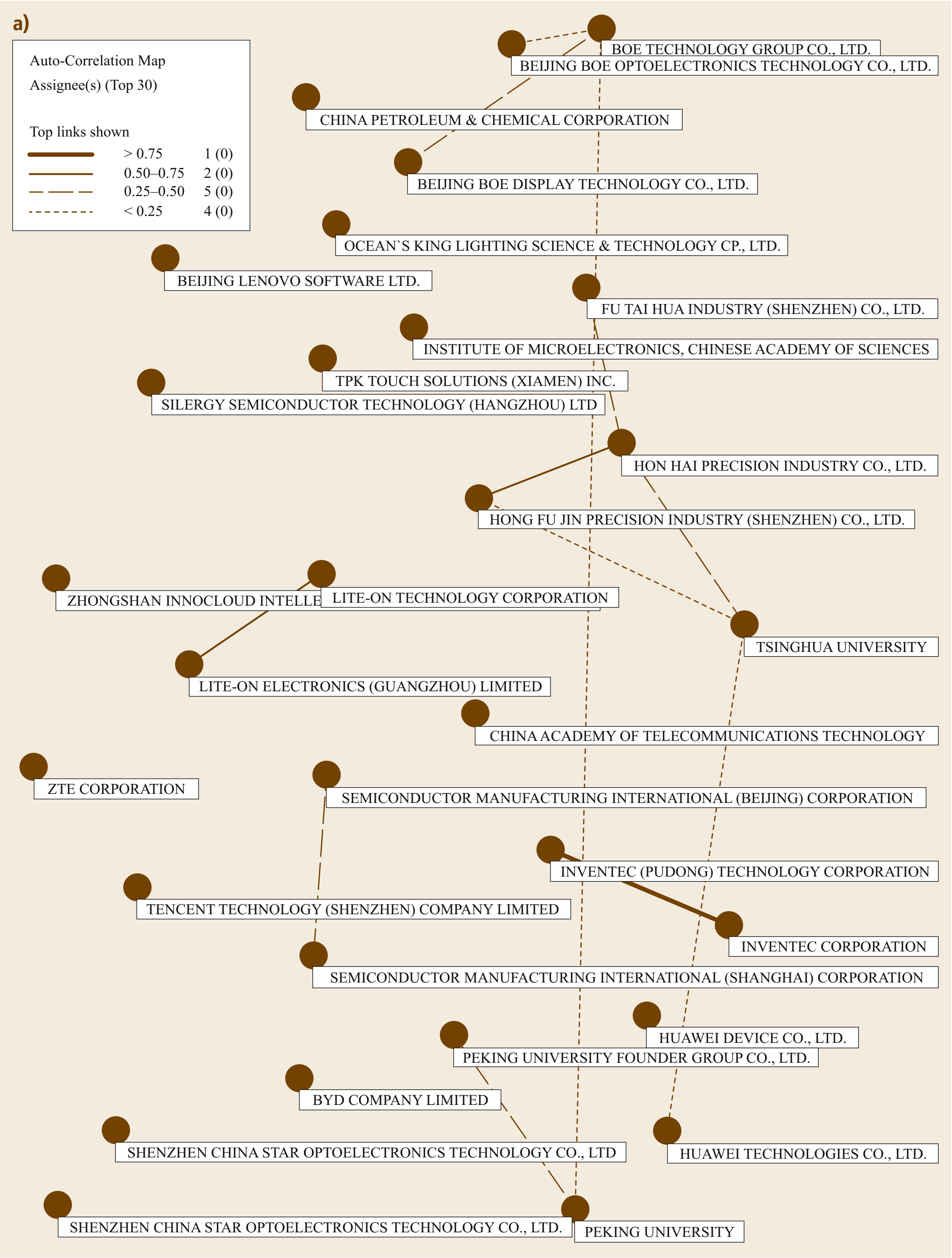


Fig. 43.8a,b University–industry network of top 30 organizations, 2015. **(a)** China; **(b)** South Korea

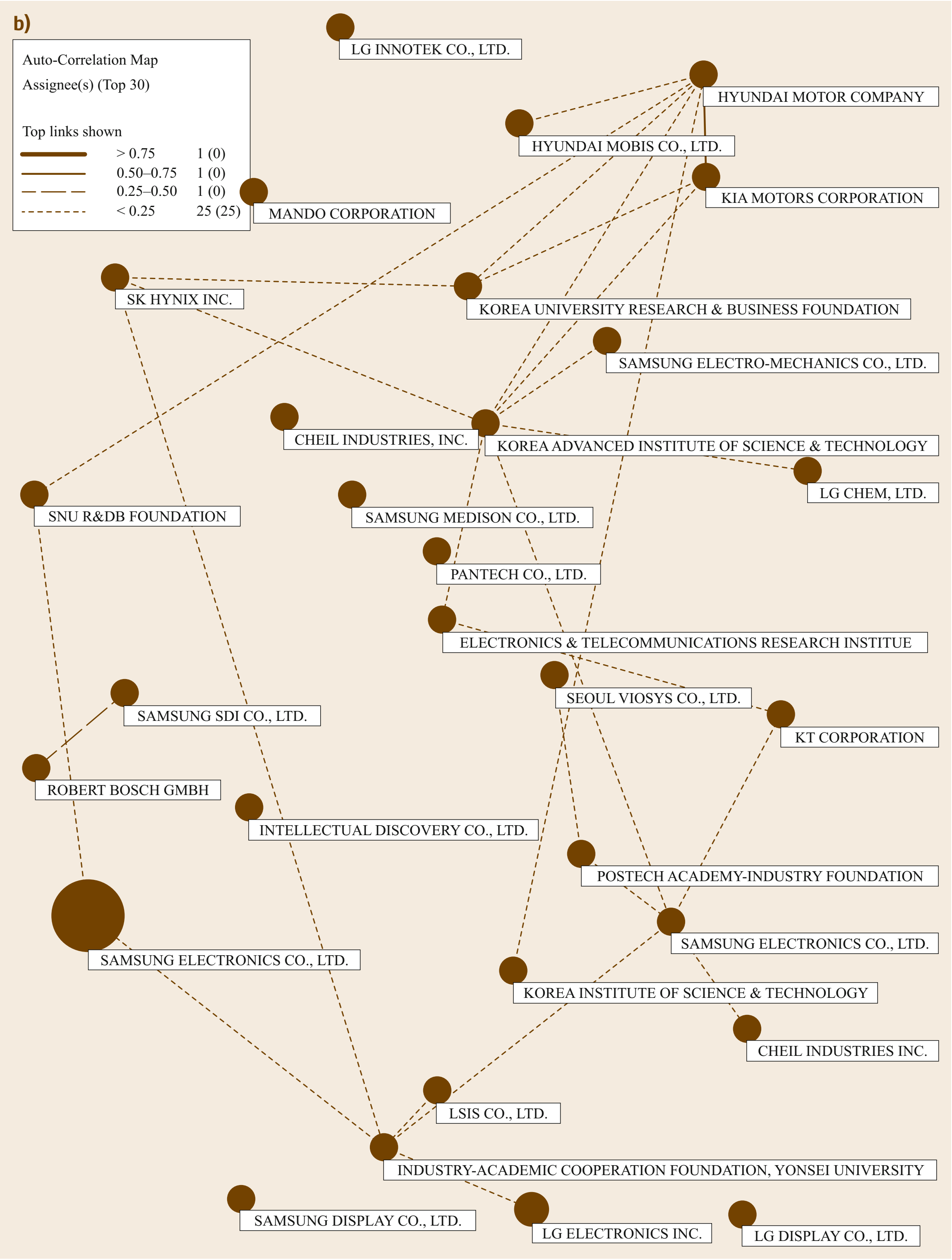


Fig. 43.8a,b (continued)

43.4 Conclusion

This chapter sought to explore the growth patterns for science, technology, and science-based technology of two economies in the course of a transition towards the postcatching-up phase. Publications and patenting data are instrumental in exploring this growth, as well as measuring elements of academic–corporate linkages that stimulate interactive learning. China and South Korea have been active in performing scientific publication and patenting activities. They witnessed tremendous impacts in terms of citations per article and a cycle of citation counts—showing both a rise and a decline. Both economies achieved a significant share of joint publications between universities and industry. South Korea in particular was able to produce a high share of application-oriented publications. The early efforts of linking science to industrial technology emerged to bear results. The links and the resultant application-oriented publications are important—particularly in the postcatching-up phase of development—as the economy sought to appropriate science-based technologies for growth.

In terms of patenting, both economies have shown the capability to produce patents and are able to converge the growth of patents with that of publications. The ratio between publications and patents for the two economies has been reduced, and our observations imply a high correlation between scientific production and technology. University–industry links for copatenting are evident and South Korea in particular had manifested a network that enables high-tech production. Technical universities such as KAIST, PRIs, and private firms featured visibly in the university–industry patenting network. This implies strong interaction between academics, researchers, and industrial engineers in terms of research activities in the context of South Korea. The network can be acknowledged as conducive for production of science-based technologies, as strong interaction between stakeholders is seen as the key factor in achieving a functional innovation system in an economy. By comparison, the patenting activities of China manifested fewer links, however, the pursuit for science-based technologies is markedly determined. Many indicators imply strong efforts towards achieving a functional innovation system.

On the basis of the cases of China and South Korea, this chapter highlighted a generic growth path for science, technology, and science-based technology in the course of a transition towards an advanced, knowledge-based economy. It is nonetheless important to explore if there are different paths pursued by other emerging economies. Therefore, a more comprehensive understanding of cyclical growth of science and technology should be sought.

References

43.1 S. Doh, B. Kim: Government support for SME innovations in the regional industries: The case of government financial support program in South Korea, Res. Policy **43**, 1557–1569 (2014)

43.2 K.-C. Cheong, C.-Y. Wong, K.-L. Goh: Technology catch-up with Chinese characteristics: What can southeast Asia learn from China?, Round Table **105**, 667–681 (2016)

43.3 C. Ling, B. Naughton: An institutionalized policy-making mechanism: China's return to techno-industrial policy, Res. Policy **45**, 2138–2152 (2016)

43.4 A. Amsden: *Asia's Next Giant: South Korea and Late Industrialization* (Oxford Univ. Press, New York 1989)

43.5 L. Kim: *Imitation to Innovation: The Dynamics of Korea's Technological Learning* (Harvard Business School Press, Boston 1997)

43.6 P.K. Wong: National innovation systems for rapid technological catch-up: An analytical framework and a comparative analysis of Korea, Taiwan and Singapore. In: *DRUID Natl. Innov. Syst. Ind. Dyn. Innov. Policy Conf., Rebild* (1999) pp. 9–12, 12 June

43.7 K. Lee, C. Lim: Technological regimes, catching-up and leapfrogging: Findings from the Korean industries, Res. Policy **30**(3), 459–483 (2001)

43.8 C.-Y. Wong, K.-L. Goh: The pathways for development: Science and technology of NIEs and selected emerging countries, Scientometrics **92**, 523–548 (2012)

43.9 C.-Y. Wong: On a path to creative destruction: Science, technology and science-based technological trajectories of Japan and South Korea, Scientometrics **96**, 323–336 (2013)

43.10 Q. Mu, K. Lee: Knowledge diffusion, market segmentation and technological catch-up: The case of the telecommunication industry in China, Res. Policy **34**, 759–783 (2005)

43.11 Y.-K. Whang, M. Hobday: Local 'test bed' market demand in the transition to leadership: The case of the Korean mobile handset industry, World Dev. **39**, 1358–1371 (2011)

43.12 H.C. Park, H.D. Hong, L. Leydesdorff: A comparision of the knowledge-based innovation systems in the economies of South Korea and the Netherlands using triple helix indicators, Scientometric **65**(1), 3–27 (2005)

43.13 J.Y. Choung, H.R. Hwang: National systems of innovation: Institutional linkages and performances in the case of Korea and Taiwan, Scientometrics **48**(3), 413–426 (2000)

43.14 L. Xue: A historical perspective of China's innovation system reform: A case study, J. Eng. Technol. Manag. **14**(1), 67–81 (1997)
43.15 B. Naughton, A. Seagal: China in search of a workable model, technology development in the new millennium. In: *Crisis and Innovation in Asian Technology*, ed. by W.W. Keller, R.J. Samuels (Cambridge Univ. Press, Cambridge 2003) pp. 160–186
43.16 C.-Y. Wong, K.-C. Cheong: Diffusion of catching-up industrialization strategies: The dynamics of east Asia's policy learning process, J. Comp. Asian Dev. **13**(3), 369–404 (2014)
43.17 S. Heilmann: Policy experimentation in China's economic rise, Stud. Comp. Int. Dev. **43**(1), 1–26 (2008)
43.18 K. Lee, J.-H. Ki: Rise of latecomers and catch-up cycles in the world steel industry, Res. Policy **46**(2), 365–375 (2016)
43.19 C.-Y. Wong, X.-S. Yap: Mapping technological innovations through patent analysis: A case study of foreign multinationals and indigenous firms in China, Scientometrics **91**, 773–787 (2012)
43.20 L. Leydesdorff, P. Zhou: Are the contributions of China and Korea upsetting the world system of science?, Scientometrics **63**, 617–630 (2005)
43.21 H.W. Park, L. Leydesdorff: Longitudinal trends in networks of university–industry–government relations in South Korea: The role of programmatic incentives, Res. Policy **39**, 640–649 (2010)
43.22 J.-Y. Choung, H.-R. Hwang: The evolutionary patterns of knowledge production in Korea, Scientometrics **94**, 629–650 (2013)
43.23 J.-Y. Choung, H.-R. Hwang, W. Song: Transitions of innovation activities in latecomer countries: An exploratory case study of South Korea, World Dev. **54**, 156–167 (2014)
43.24 J.A.T. da Silva: Does China need to rethink its metrics- and citation-based research rewards policies?, Scientometrics **113**, 1229–1231 (2017)
43.25 P. Zhou, L. Leydesdorff: The emergence of China as a leading nation in science, Res. Policy **35**(1), 83–104 (2006)
43.26 K. Lee, F. Malerba: Catch-up cycles and changes in industrial leadership: Windows of opportunity and responses of firms and countries in the evolution of sectoral systems, Res. Policy **46**(2), 338–351 (2017)
43.27 P.S. Meyer: Bi-logistic growth, Technol. Forecast. Soc. Change **47**, 89–102 (1994)
43.28 C.Y. Wong, K.L. Goh: Growth behavior of publications and patents: A comparative study on selected asian economies, J. Informetrics **4**(4), 460–474 (2010)
43.29 A.T. Bernadas, E.M. Albuquerque: Cross-over, thresholds, and interactions between science and technology: Lessons for less-developed countries, Res. Policy **32**(5), 865–885 (2003)
43.30 U. Schmoch: Double-boom cycles and the comeback of science-push and market-pull, Res. Policy **36**, 1000–1015 (2007)
43.31 R. Mazzoleni, R.R. Nelson: Public research institutions and economic catch-up, Res. Policy **36**, 1512–1528 (2007)
43.32 C.-Y. Wong, M.-C. Hu, J.-W. Shiu: Collaboration between public research institutes and universities: A study of industrial technology research institute, Taiwan, Sci. Technol. Soc. **20**, 161–181 (2015)
43.33 R. Stankiewics: Technology as an autonomous, socio-cognitive system. In: *Dynamics of Science Based Innovation*, ed. by H. Grupp (Springer, Berlin 1992) pp. 19–44
43.34 H. Grupp: Spillover effects and the science-based of innovations reconsidered: An empirical approach, J. Evolut. Econ. **6**(2), 175–197 (1996)
43.35 H.M. Jarvenpaa, S.J. Makinen, M. Seppanen: Patent and publishing activity sequence over a technology's life cycle, Technol. Forecast. Soc. Change **78**(2), 283–293 (2011)
43.36 L. Colledge, R. Verlinde: *Scival Metrics Guidebook* (Elsevier, Amsterdam 2014)
43.37 R.J.W. Tijssen: Discarding the 'basic science/applied science' dichotomy: A knowledge utilization triangle classification system of research journals, JASIST **61**, 1842–1852 (2010)
43.38 B. Van Looy, C. Vereyen, U. Schmoch: Patent statistics: Concordance IPC v8—NACE rev 2. Eurostat, https://circabc.europa.eu/sd/a/d1475596-1568-408a-9191-426629047e31/2014-10-16-Final%20IPC_NACE2_2014.pdf (2014)
43.39 C.-Y. Wong, H.-N. Fung: Science-technology-industry correlative indicators for policy targeting on emerging technologies: Exploring the core competencies and promising industries of aspirant economies, Scientometrics **111**(2), 841–867 (2017)
43.40 OECD: ISIC Rev. 3 Technology Intensity Definition. OECD Directorate for Science, Technology and Industry Economic Analysis and Statistics Division, https://www.oecd.org/sti/ind/48350231.pdf (2011)
43.41 H. Grupp: *Foundations of the Economics of Innovation, Theory, Measurement and Practice* (Edward Elgar, Cheltenham 1998)
43.42 U. Schmoch: *Concept of a Technology Classification for Country Comparisons. Final Report to the World Intellectial Property Office (WIPO)* (Fraunhofer ISI, Karslruhe 2008)
43.43 J.-S. Shin: Dynamic catch-up strategy, capability expansion and changing windows of opportunity in the memory industry, Res. Policy **46**(2), 404–416 (2016)
43.44 L. Bornmann, L. Leydesdorff: Scientometrics in a changing research landscape, EMBO Rep. **15**, 1228–1232 (2014)
43.45 K.-S. Kwon, H.-W. Park, M. So, L. Leydesdorff: Has globalization strengthened South Korea's national research system? National and international dynamics of the triple helix of scientific co-authorship relationships in South Korea, Scientometrics **90**, 163–176 (2012)
43.46 K. Lee: *Schumpeterian Analysis of Economic Catch-up: Knowledge, Path-Creation and the Middle-Income Trap* (Cambridge Univ. Press, Cambridge 2013)
43.47 K. Lee: *Economic Catch-up and Technological Leapfrogging: The Path to Development and Macroeconomic Stability in Korea* (Edward Elgar, Cheltenham 2016)

Chan-Yuan Wong

Institute of Technology Management
National Tsing Hua University
Hsinchu City, Taiwan
wong_chanyuan@um.edu.my

Chan-Yuan Wong is a faculty member at the Institute of Technology Management, National Tsing Hua University of Taiwan. He was a Senior Lecturer at University of Malaya and also consultant for Malaysia's Ministry of Science,Technology and Innovation (MOSTI) and Malaysian Industry-Government Group for High Technology (MIGHT) for several research projects, and IDRC-sponsored projects on Innovation Policies and Inclusive Development of Asia. He has published on industrial and innovation systems, public policies, R&D management and growth models for production of science and technology.

Hon-Ngen Fung

Post-graduate and Research Centre
KDU University College
Shah Alam, Malaysia
kennethfung84@gmail.com

Hon-Ngen Fung is a Lecturer at the KDU University College. His research interests focus on scientometrics, innovation systems, and science and technology policy. He has published in several international journals including *Scientometrics*, *TechnologicalForecasting and Social Change*, and*Social Science Information*.

44. Standardization and Standards as Science and Innovation Indicators

Knut Blind

The focus of innovation policies has shifted from knowledge creation and protection (e. g., by patents) to knowledge diffusion (e. g., via open access) in order to promote their implementation. This has led to an increasing need for innovation indicators that reflect the implementation of knowledge within innovative products and services. Standardization as a kind of open innovation process, and standards as its output, represents a new type of innovation indicator. In this chapter, we begin with a discussion of existing opportunities for using standards and standardization as innovation indicators, including three specific examples of input, throughput, and output indicators. Next we identify challenges that must be addressed to close the data gaps—which are still very significant when compared with patent data. In addition, the broader concept of quality infrastructure is introduced in order to point out the complexity of standards implementation, and its close link to innovation as well. The chapter concludes with examples of how decision makers in industry and policy could make use of a comprehensive database of standardization and standards to evaluate innovation policy initiatives.

44.1 Background

In the 1990s, the increasing relevance of intellectual property rights (IPR), particularly with regard to patents and the accessibility of patent data via resources such as PATSTAT (Worldwide Patent Statistical Database), provided by the European Patent Office and the OECD (Organisation for Economic Co-operation and Development), led to a virtuous cycle accompanied by the broadening of areas of investigation and improved data quality. However, despite the increasing relevance of IPR, the number of patent applications, especially in Western countries, has been stagnating or even decreasing [44.1]. In parallel, the increasing opportunities and demands for measuring the scientific productivity of countries, research institutes, and even individual researchers has led to the development of bibliometrics as a new discipline based on scientific publications and related information (see other contributions in the handbook). Recently, however, the importance of accessing, disseminating, and implementing knowledge—in contrast to knowledge creation and protection—has been receiving greater attention from policymakers. One example is the shift in focus of the current European Framework Programme Horizon 2020 from research to innovation, and fostering open access to scientific data and open innovation [44.2, 3].

Methods for measuring the implementation of innovative ideas have been addressed by researchers focusing on patents, but the question of whether patents are actually used can only be answered by inventor surveys, such as those recently reported by *Torrisi* et al. [44.4] and *Ploschka* [44.5], or by specific case studies, for example, focusing on complex products such as laptops [44.6]. Certainly, analyzing innovative ideas that have actually been implemented in new products is the more accurate, but much more cumbersome approach. An alternative and complementary approach is the analysis of the development and implementation of standards, as the additional effort to develop common standards is a strong indication that the innovative ideas are eventually implemented in products and processes, and are not used only for strategic purposes, such as the use of patents to block competitors or

W. Glänzel et al. (Eds.), *Springer Handbook of Science and Technology Indicators*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-02511-3_44

in negotiations [44.4, 7]. Examples of the most recent standardization activities include topics such as cybersecurity, blockchain, fifth-generation wireless systems (5G), the Internet of Things (IoT), cloud and big data, additive manufacturing, artificial intelligence, robotics, and augmented/virtual reality. Furthermore, standards also include product and process characteristics (e. g., *Lorenz* et al. [44.8] in the case of biotechnology), which is less the case for patents [44.9]. Finally, standards not only address innovative technological aspects, but increasingly encompass services [44.10, 11]. These include online gaming and management systems [44.12], focusing particularly on quality management [44.13], and recently extending to issues such as environmental management [44.14] and corporate social responsibility [44.15]. In contrast to patenting, which takes place within closed inventor teams, and in addition to open-source activities focused specifically on software [44.16], standardization in publicly accredited national, European, or international bodies and informal consortia, particularly those active in information and communication technology (ICT) [44.17] and increasingly in multimodal form [44.18], is serving to generate common knowledge, which is especially interesting for the knowledge sourcing of small and medium-sized enterprises [44.19]. In summary, standards and standardization provide opportunities to generate innovation indicators that not only cover a much broader spectrum of innovation, but also reflect dynamic and open innovation processes via the development of common standards, their revision, and eventual replacement by the next generation of standards [44.20].

44.2 Definitions and Processes

The regulatory framework generally comprises regulations released and enforced by governmental institutions. Industry and other affected stakeholders may complement these governmental regulations by self-regulatory coordination [44.21]. Their efforts can result in voluntary commitments and standards released by publicly accredited or even administrative standardization bodies.

Standardization is defined according to European standard EN 45020:2006 as the

> activity of establishing, with regard to actual or potential problems, provision for common and repeated use, aimed at the achievement of the optimum degree of order in a given context. [44.22]

The benefits of standards are generally differentiated as follows: securing interoperability, especially in network industries; assuring quality, health, and safety; and the generic functions of both reducing variety and enabling the exploitation of economies of scale and information sources such as scientific publications and patents [44.23, 24]. Given its tried-and-true processes, standardization enjoys a high degree of legitimacy and is uncontested in terms of antitrust legislation.

The result of a standardization process is a standard, defined according to EN 45020:2006 as

> a document established by consensus and approved by a recognized body, that provides, for common and repeated use, rules, guidelines or characteristics for activities or their results, aimed at the achievement of the optimum degree of order in a given context.

recently confirmed by OECD/Eurostat [44.25].

Standards are the result of standardization work at the national, regional (European), and international levels. Anyone can submit a proposal for a new standard. Standards are drafted by committees set up by national standardization bodies, the European Committee for Standardization (CEN) and the European Committee for Electrotechnical Standardization (CENELEC), or, on the international level, the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), working in agreement with defined principles and rules of procedure and design. Any party interested in the drafting of standards is able to participate in the work of standardization bodies. The term *interested party* refers to such groups as manufacturers, consumers, retail businesses, science or research institutes, insurance companies, governmental authorities, or testing institutes dispatching experts to the working body or one of its specialist areas. National interests are represented in CEN/CENELEC and ISO/IEC by experts and delegations from the national standardization bodies. Their staff takes care of coordinating standardization work at the national, European, and international levels. Standards are reviewed for relevance at least once every five years. If a standard is shown to fall short of the state of the art, its content will be revised or the standard will be withdrawn altogether.

There are various ways for experts from businesses and organizations to become involved in standardization work (Fig. 44.1). The type of involvement and the amount of effort required depend on the interests and available resources that stakeholders may have.

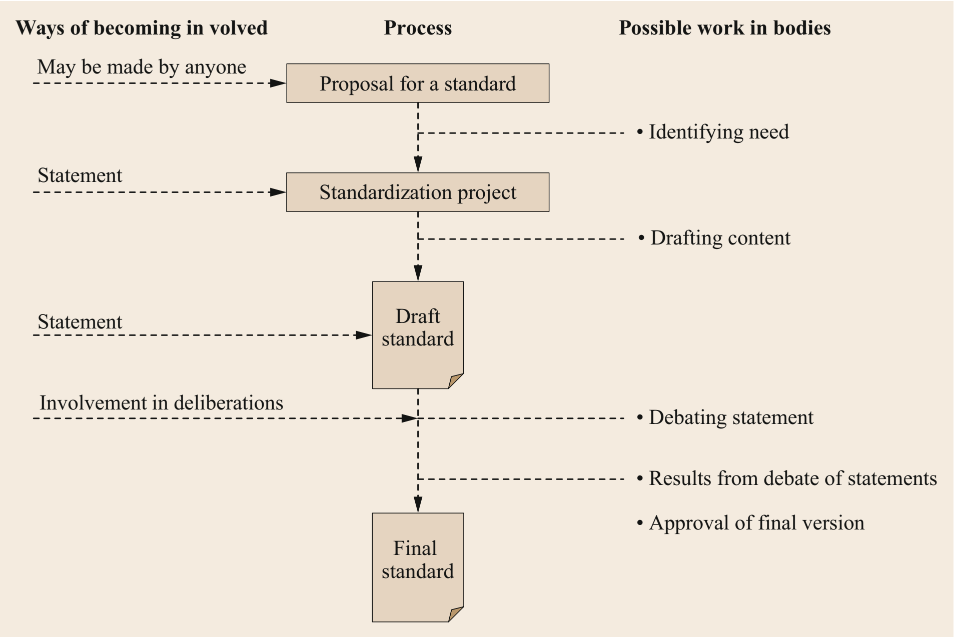


Fig. 44.1 The standardization process based on [44.26]

44.3 Current Opportunities

Standards contain knowledge that is relevant not only for the production of products and services, but also for all interested stakeholders active in the research [44.27, 28] and development [44.8, 29] process. Examples include both the terminology and the measurement and testing standards facilitating communication within research and development, and also health, environmental quality, and safety standards necessary to create trust among early adopters of innovative new products and services, in addition to compatibility standards as the basis for generating positive network externalities [44.30].

Standardization can also be linked to both scientific publications [44.31, 32] and patenting. Thus far, only standard-essential patents with higher value as measured by forward citations have been the subject of intensive empirical investigations, starting with the seminal contribution of *Rysman* and *Simcoe* [44.33] based on linking patent databases with information provided by the standards development organizations (see the most recent review by *Pohlmann* and *Blind* [44.34]). The work in this field has been further extended to the analysis of motivations and strategies for including patents in standards [44.35–38] and their impact on innovation [44.20] or company performance [44.39, 40].

In parallel, company participation in standardization bodies [44.11, 41, 42] and consortia [44.43] has been studied extensively based on either surveys or membership information [44.44], and confirms the generally positive relationship between innovation activities and success on the one hand, and standardization on the other. However, in contrast to the large body of literature on individual inventors, individual researchers' involvement in standardization is a rather unexplored topic, as the information is largely confidential and can only be accessed via internal databases [44.31] or surveys [44.5, 45]. Here, the relationship to researcher performance measured by scientific publications or patent applications yields ambivalent results [44.31], whereas the inclusion of publications in standards—as in the case for standard-essential patents—generates both more and longer-lasting citations (e. g., for biotechnology *Raven* and *Blind* [44.32]).

In addition to standards published by standardization bodies or consortia, numerous company standards are developed, particularly by large companies, both for internal use and for coordinating their supply chains [44.46], which according to *De Vries* [44.47] might even outweigh the number of applied formal standards. However, these documents are either confidential, such as trade secrets, or available only to suppliers. Therefore, they are not suitable as a basis for constructing representative and valid indicators, although they do codify internal company knowledge, processes, and preferences that may complement inventions, patents, or even trade secrets.

On a macroeconomic level, similar to the case with patents, a generally positive association has been found for the impact of standards as indicators of countries' or industries' innovativeness, measured as stocks of standards provided by a commercial database operated by standards development organizations, on economic growth [44.48], trade [44.49], and global value chains [44.50]. While the number of existing standards [44.51] is a rather imprecise measure, the International Organization for Standardization (ISO) provides data about the implementation of the most important management standards, such as the more than one million certifications for the ISO 9001 quality management system [44.52] and the ISO 14001 environmental management system. However, whereas the relationship between the ISO 9001 certification and innovation is uncertain on the company level [44.53], the link between ISO 14001 and innovation is positive, at least at the country level [44.54]. In contrast to the dominance of Western countries in providing input into standardization processes, data regarding the implementation of these management standards is available for all of the more than 150 member countries of ISO, allowing for global analyses. Recently, *Ehrich* and *Mangelsdorf* [44.55] analyzed the impact of certifications related to private standards on trade flows in the food sectors, again with a focus on developing countries.

Unfortunately, systematic data regarding the implementation of other standards is not available. A first study focused on the number of standards implemented, for example, in the production of a laptop [44.6]. In a recent work, *Ploschka* [44.56] identified the implementation of standards within the whole product portfolio of a company manufacturing hydraulic products. Although this approach requires significant effort, this study confirms at least the feasibility of such an approach.

Nevertheless, the options for data collection have since been expanded within country-wide company surveys. For example, previous British editions of the Community Innovation Survey included questions about the role of standards as innovation sources and barriers [44.57, 58], and German editions of the community innovation survey have explicitly distinguished between standards and regulations as barriers to innovation on an ongoing basis, thus enabling longitudinal analyses [44.59]. A recent survey of the active involvement of German companies in standardization [44.60] revealed that 10% are actively involved—a share similar to that for patenting. In fact, the majority of companies are following both strategies. In addition, data provided by national standardization offices are matched to the community innovation survey, for example, in the Netherlands [44.11] and Germany [44.61]. Together, these studies show positive correlations between innovation activities and success on the one hand, and the role of standards as information sources or participation in standardization on the other.

In 2012, the German Standardization Panel was launched with a pilot survey addressing the 10 000 companies active in the national standardization body DIN e.V. Since then, seven annual panel surveys have been conducted and have included more than 1000 companies constituting over 10% of the universe of standardizing companies. Efforts are currently under way to involve the more than 100 000 German companies purchasing standards, assuming that they implement standards without being actively involved in standardization [44.62]. With the fifth wave of the panel starting in October 2016, the German Ministry of Economic Affairs and Energy (BMWi) has taken over patronage for the envisaged long-term initiative. Research and other institutions such as test laboratories are an additional group targeted for surveys, in order to complement the business perspective (see *Blind* et al. [44.50] for a pilot study focusing on a research institute).

Examples

To illustrate the actual opportunities for using standards as innovation indicators, we present three examples, representing input, throughput, and output indicators.

According to the Commission of Experts for Research and Innovation (EFI) [44.63], standardization is an important factor in the commercialization of innovative technologies. Since we have no internationally comparable information about efforts invested in standardization, we must rely on data provided by the International Organization for Standardization (ISO). By participating in the committees at ISO, a country can make a significant impact on global technical infrastructures. German companies are more frequently involved in the work of the ISO than companies of any other country, as measured by the number of German secretariats (Fig. 44.2).

The more than 3500 committees at ISO have formulated and published over 20 000 international standards and standards-type documents, totaling almost one million pages. The largest share of documents, at 27.3%, covers engineering technologies, while 21.8% relate to material technologies (Tab. 44.1). The other half of the ISO standards include electronics and information technology and telecommunications, at 17.7%, and another 10.7% involve the transportation of goods. A cross-sectional category including generalities, infrastructures, sciences, and services is responsible for 9.3%. Agriculture and food technologies constitute

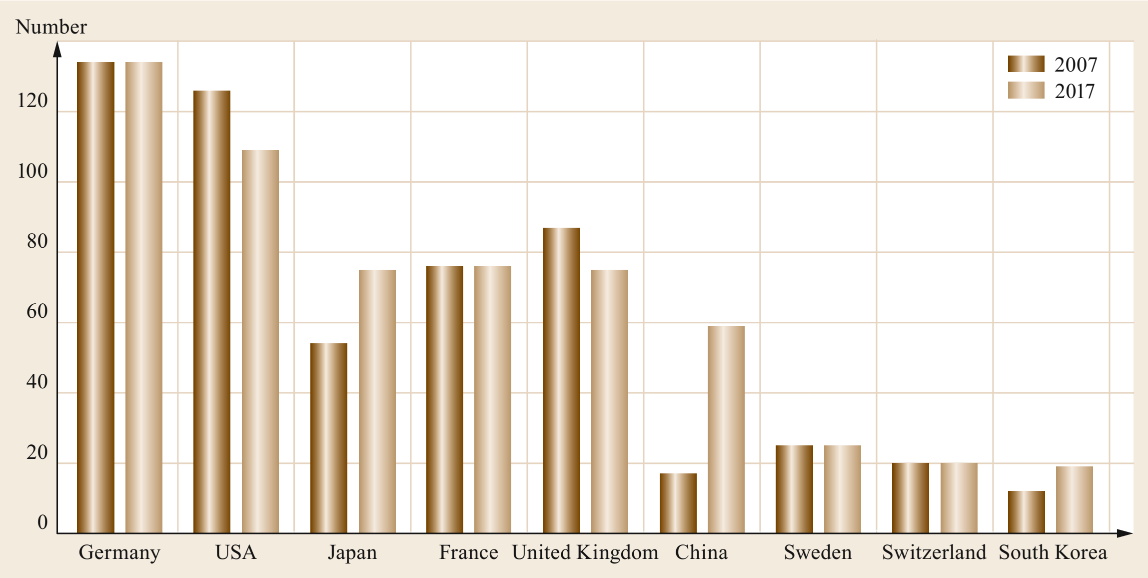


Fig. 44.2 Number of secretariats listed by the technical committees and subcommittees of the International Organization for Standardization (ISO). Source: Own diagram based on ISO (2007:15) and [44.64]. ©EFI-Commission of Experts for Research and Innovation 2017

5.6% of ISO standards, and 2.5% focus on construction. Another important set of standards, focused mainly on health, safety, and environmental protection, represents 4.1%. Finally, 1% of the ISO standards deal with special technologies, including new areas. Information regarding the ISO standards can be accessed through the ISO homepage. Professional database providers, including national standardization bodies, offer bibliographic data regarding standards released by the majority of ISO members, primarily as a resource to support industry with their standards and technical regulation inquiries; these are only partially useful for scientific purposes such as the construction of innovation indicators.

Most of the 166 national members of ISO include the stock of ISO standards in their national stock of standards, especially since few members still produce a significant number of national standards. However, the integration of ISO standards reflects the economic capacity, and thus the capacity for innovation, of the various countries, as national industries are not interested in international standards related to technologies in which they are not active. Finally, like patents, the stocks of standards can be considered throughput indicators, because the publication of standards does not guarantee their implementation. However, the standardization process—in contrast to patenting—requires the exchange and even the common production of knowledge between the participants, generating an economic impact even before implementation.

Table 44.1 Shares of ISO standards by technical sector as of the end of 2016 [44.65]

Technical sector	Shares (%)
Engineering technologies	27.3
Material technologies	21.8
Electronics, information technology and telecommunications	17.7
Transport and distribution of goods	10.7
Generalities, infrastructures, sciences, and services	9.3
Agriculture and food technology	5.6
Health, safety, and the environment	4.1
Construction	2.5
Special technologies	1

In addition to the more than 20 000 ISO standards, the International Electrotechnical Commission (IEC) provides nearly 10 000 international standards focusing on electrotechnology, and the International Telecommunication Union (ITU) maintains a portfolio of around 5000 standards. In Europe, the three standardization organizations—the European Committee for Standardization (CEN), the European Committee for Electrotechnical Standardization (CENELEC), and the European Telecommunications Standards Institute (ETSI)—are responsible for almost 30 000 standards, which include the majority of ISO standards.

Table 44.2 shows the stock of national standards including international and—for the European countries—European standards for a selected sample of

Table 44.2 New standards, withdrawn standards, and total stock of standards in selected countries at the end of 2017 (Source: Based on [44.66])

Country	New documents per month	Present stock	Withdrawn documents with substitution per month	Withdrawn documents without substitution per month
Germany	380	39 575	367	20
United Kingdom	573	41 452	177	379
Russia	495	33 310	11	12
Austria	319	32 564	199	11
China	301	34 482	89	0
South Korea	296	20 250	232	9
France	235	38 903	88	171
Spain	151	31 643	110	13
United States (ANSI)	136	11 155	92	14
Poland	118	29 584	59	19
Sweden	107	24 047	98	20
South Africa	64	7556	15	18
Brazil	56	8744	35	11

countries. We see that most European countries provide European and international standards, and only large countries, namely Germany, the United Kingdom, and France, still maintain a significant number of national standards. In the United States, there are several sector-specific standardization organizations that provide additional standards. Therefore, the total number of standards is much higher than those administered by the American National Standards Institute (ANSI). Emerging countries have a significantly lower, but growing, number of standards. There is also a significant dynamic in the stock of standards: 10% are newly published each year, and almost the same share are withdrawn. Thus, over a period of 10 years, there is a complete turnover in the stock of standards.

Finally, the implementation of standards can be considered an output indicator. However, as with patent implementation, the challenge lies in the lack of a database. The only internationally comparable database regarding the implementation of standards is the ISO survey of management system standard certifications [44.67]. The *ISO Survey* is an annual survey of the number of valid certifications of ISO management standards worldwide (Tab. 44.3). To compile this information, accredited (i. e., independently evaluated) certification bodies are asked each year to provide information about the number of valid certificates that they have issued (ISO itself does not perform certification and therefore does not issue certificates). This results in the most comprehensive overview of certifications to these standards currently available, notwithstanding the fluctuations in the number of certificates from year to year due to differences in the number of participating certification bodies and the number of certificates they report. In particular, the more than one million certifications of ISO 9001 quality management standards and the over 300 000 environmental management standards certifications enable the construction of indicators by country, which over time reflect the diffusion of management innovations.

44.4 Future Challenges

In our critical evaluation of existing opportunities for using standardization activities and published standards as research and innovation indicators, we observe significant fragmentation in various dimensions. Consequently, a comprehensive approach is needed.

In contrast to the basic Community Innovation Survey that is mandatory for all Member States of the European Union, standardization has been included in the national editions only in the United Kingdom and Germany. In addition, the questionnaire versions are not even aligned between these two countries, and consistency over time, i. e., between different waves, has not been achieved within the British survey. Specific efforts in setting up panels wholly dedicated to standardization have been implemented only in Germany so far. No attempts have yet been made to expand this nationally focused initiative to other countries or to the European or even international level.

Table 44.3 Number of certificates issued related to the most important ISO standards [44.67]

Standard	Number of certificates in 2015
ISO 9001 (Quality management system (QMS))	1 033 936
ISO 14001 (Environmental management)	319 324
ISO 50001 (Energy management)	11 985
ISO 27001 (Information security management)	27 536
ISO 22000 (Food safety management)	32 061
ISO/TS 16949 (Quality management system for the automotive industry)	62 944
ISO 13485 (Medical device QMS)	26 255
ISO 22301 (Business continuity management)	3133
ISO 20000-1 (IT service management systems)	2778
Total	1 519 952

In addition, whereas PATSTAT provides patent data in a standardized format for all major patent offices, a similar database is not available for standards. High-quality data, while available for a few European countries, does not exist for the majority of European countries or the OECD member states. Furthermore, the rather decentralized standardization system in the United States, with several hundred standardization bodies, represents a fragmented landscape itself, which poses a challenge for country-specific approaches in Europe. A review by *Zoo* et al. [44.68] revealed that quantitative studies on standards in emerging economies are rare, and those that do exist focus primarily on China.

To effectively address these two dimensions of fragmentation, both aspects must be considered. On the one hand, the topic of standardization would need to be included in the basic Community Innovation Survey module required for all member states, on a regular cycle of every 4 or even 2 years. Alternatively, an international standardization survey could be established, as we observe a shift in standardization activities not only from the national to the European level, but to the international level as well, driven by the numerous standardization consortia active at the international level. On the other hand, the fragmentation and heterogeneity in quality related to standards databases would require a joint international initiative similar to those achieved by PATSTAT.

In addition, two further challenges that have not yet been successfully tackled by the Community Innovation Survey or existing patent databases should be noted. As we know from the vast amount of previous research, close interaction between the various stakeholders and institutions is needed to realize a well-functioning innovation system [44.69]. However, the community innovation survey is focused on businesses and their perspectives. This narrow focus must be widened to include both research institutes and other institutions on the one hand, and consumers on the other. Such an expansion would also be beneficial for standardization surveys, given the potentially important contributions of research institutes to the development of standards already validated by research on standard-essential patents and publications, and the high relevance of users for their implementation. However, public sector stakeholders must be integrated as well, as they contribute to standardization, for example, in aligning standards with the regulatory framework, and they must rely on standards, as in the case of public procurement [44.27].

As already mentioned, existing patent databases provide information only on the application details and current legal status of a patent, and not its actual implementation in products and services. Although we can assume that stakeholders would not invest resources in a standardization process unless they were interested in implementing the standard, as there would be little strategic motivation (e. g., blocking competitors via standards) [44.19], we face a similar measurement challenge related to standards. Standards databases generally do not provide information on their implementation, and the exception of the *ISO survey* [44.67] focuses on the few very visible management standards relying on certification data. However, this type of information would provide valuable insight into the dissemination of standards. This becomes even more relevant by considering that developing countries benefit mainly from implementing and not from developing standards, as was recently argued and confirmed by *Zoo* et al. [44.68].

As discussed in an earlier work by *Blind* [44.70], the focus on standardization and standards as innovation activity and results must be widened to encompass the whole quality system, a term first mentioned by *Guasch* et al. [44.71], or the whole quality infrastructure, a term introduced by *Sanetra* and *Marbán* [44.72]. A quality infrastructure is a system of institutions that jointly ensure that products, processes, and services meet certain predefined specifications. Most notably, it enables companies to improve both their production processes and their products, facilitating compliance with regulations or international requirements, especially international standards. The elements of quality infrastructure, in addition to standardization and the standards created, include conformity assessment (i. e., certification), accreditation and market surveillance, as well as metrology. These components are highly complementary to standards, as well as to technical regulations recently introduced by the OECD, in partnership with other organizations [44.73], for countries' analysis of policies targeting small and medium-sized companies. As an example, it is not possible to formulate a standard regarding the properties of a product unless the respective measurement units are defined, and adequately calibrated measurement instruments exist. The standard is effective only if producers comply with it, which is more likely if there are incentives to do so—for instance, if compliance could be signaled to potential consumers through a certification scheme. The impacts of certification increase in significance in proportion to the trust that customers have in the certification institute, which is higher if an accreditation body has evaluated the competence of the certifier (e. g., *Blind* et al. [44.74]). The high interdependence between the various elements of the quality infrastructure requires a comprehensive or holistic approach in which these elements and their activities are perceived as integral parts of a complex system (for a first economic framework see [44.58]). They cannot be tackled as independent activities isolated from the functioning of the whole. This is an important consideration when setting up a new quality infrastructure or improving an existing one. It requires an understanding of how the components support each other, the embedding of each element within the whole system, and the framing of this infrastructure as a pillar of the national economy or innovation system [44.75].

The overall availability of data regarding standardization activities or standards (see *Blind* [44.76, 77] for surveys) and their impact on—or more appropriately, their interrelation with—innovation [44.78] and companies' commercial success must now catch up to the already available data on research and innovation activities on the one hand, and their successes and impacts on the other. In addition, going one step further requires including other innovation system actors in such survey approaches, as well as addressing the implementation of standards within the national quality infrastructure to better capture their effective dissemination over time, countries, and industries.

44.5 Relevance for Decision Makers in Industry and Policy

These new opportunities to collect data regarding standards and their implementation could be used by decision makers in both industry and policy.

Recent studies (e. g., *Großmann* et al. [44.29]) reveal that, despite the relevance of patents and standards for new product development, they have not yet been integrated into strategic (technology) management. Screening of the existing standardization landscape and stocks of standards, however, might inform companies' decision makers of opportunities to set their own standards, as well as the need to implement existing standards. Recent reviews of both traditional technology transfer channels [44.79] and open innovation [44.80, 81] do not include the option of standardization, even though knowledge sourcing in standardization processes is particularly relevant for small companies [44.19]. Clearly, opportunities for standardization are far from being effectively exploited by industry.

From a policymaker perspective, available indicators related to both standardization and standards can be used to evaluate existing research and innovation programs and to design new programs, as well as to assess a country's innovation policy and system in general. One example is the Lead Market Initiative (LMI) established by the European Commission [44.82], which focused primarily on the three instruments of public procurement, legislation, and standardization [44.83]. The final evaluation of the LMI [44.84] could have been based on rigorous quantitative analyses using time-series data, even allowing the application of a difference-in-differences approach. The European research and innovation program Horizon 2020 that has been ongoing since 2014 also includes standardization processes and standards, in addition to scientific publications and patents, as success indicators for innovation [44.85–87]. Here again, control groups of companies not being promoted by the program and previously not involved in standardization could be generated based on the Community Innovation Survey, including sections on companies' involvement in standardization and implementation of standards.

As discussed in the section on examples, the Commission of Experts for Research and Innovation (EFI), which advises the German federal government, has included the number of secretaries within ISO by country as a performance indicator, in addition to scientific publications and patents, in their annual report on research, innovation, and technological performance in Germany [44.63]. Since the dominant position of both Germany and the United States is challenged by the expanding international standardization activities of China and Brazil as emerging countries, future innovation policy initiatives might consider subsidizing companies' participation in international standardization committees to represent national interests, both in a commercial sense and in ethics debates related to genetically modified organisms or privacy issues such as data protection in the cloud [44.88].

Furthermore, standards are an element of the national regulatory framework [44.3], and constitute a technical barrier to trade (TBT) at the international level when standards differ between countries. This was a critical issue, for example, during the now abandoned negotiations on the Transatlantic Trade and Investment Partnership (TTIP) [44.89, 90]. Since institutions are important elements of national innovation systems, institutional changes have implications for the level and direction of innovation. Consequently, historical data regarding the impact of standards on innovation could provide a sound database for an ex ante assessment of the effects of institutional changes as a result of standardization systems in the European Union and its Member States on future innovations (see *Blind* [44.76, 77] for current insight into the impact of standardization and standards on innovation). In this context, it should be mentioned that standardization bodies are increasingly performing foresight analysis in order to anticipate future challenges, enabling a proactive rather than reactive approach [44.91].

In summary, the policy relevance of standardization in relation to innovation is evolving and progressing from the OECD members to the global level, especially in the large and fast-growing emerging countries. Yet, despite the launch of the LMI by the European Commission 10 years ago [44.82] and the recent incorporation of benchmarking measures focusing on small and medium-sized companies into OECD policy [44.73], the potential for standardization as an effective and efficient instrument for promoting innovation is still not fully exploited by policymakers. For example, in the European Commission's new publication [44.3] on open innovation, open science, and openness to the world, in addition to the value of international collaboration in science and research as a means of fostering open innovation to improve scientific quality and research results, the importance of global standard-setting is discussed as a way to more effectively tackle global challenges, facilitating participation in global value chains and new and emerging markets. However, standards are not explicitly considered in regulatory reform measures, for example, in the implementation of strategies to promote open innovation. Similarly, in efforts to promote open science, standards are not accepted as another output of research and innovation activities in addition to publications and patents, although this was addressed in Horizon 2020 [44.87]. Because the potential for standardization as a strategic management and policy instrument has not been effectively exploited, the related use of standardization and standards as science and technology indicators is still limited. However, the opportunities described may serve to drive the integration of standards into the portfolio of existing and well-established science and technology indicators.

Acknowledgments. A previous shorter version of this paper was presented at the OECD Blue Sky Forum on Science and Innovation Indicators in Ghent in September 2016. The author thanks two anonymous reviewers for their comments, which helped me to improve the chapter significantly. Funding has been received from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 778420—EURITO.

References

44.1 WIPO: *World Intellectual Property Report Breakthrough Innovation and Economic Growth* (WIPO, Geneva 2015)

44.2 C. Moedas: Open innovation, open science, open to the world. In: *A New Start for Europe: Opening Up to an ERA of Innovation, Conference Speech/15/5243* (European Commission, Brussels 2015)

44.3 European Commission: *Open Innovation, Open Science, Open to the World: A Vision for Europe* (European Commission, Brussels 2016)

44.4 S. Torrisi, A. Gambardella, P. Giuria, D. Harhoff, K. Hoisl, M. Mariani: Used, blocking and sleeping patents: Empirical evidence from a large-scale inventor survey, Res. Policy **45**(7), 1374–1385 (2016)

44.5 A. Ploschka: *R&D, Patents and Standards: Three Studies on the Application of Standards, the Involvement in Standardization and the Relation to Innovation*, Ph.D. Thesis (Technical Univ., Berlin 2016)

44.6 B. Biddle, A. White, S. Woods: How many standards in a laptop? (And other empirical questions). In: *Proc. 2010 ITU-T Kaleidosc. Acad. Conf.*, Vol. 5682128 (2010) pp. 123–130, http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5682128

44.7 K. Blind, J. Edler, R. Frietsch, U. Schmoch: Motives to patent: Empirical evidence from Germany, Res. Policy **35**(5), 655–672 (2006)

44.8 A. Lorenz, M. Raven, K. Blind: The role of standardization at the link between product and process development in biotechnology, J. Technol. Transf. (2017), https://doi.org/10.1007/s10961-017-9644-2
44.9 W. Cohen, R. Nelson, J. Walsh: *Protecting Their Intellectual Assets: Appropriability Conditions and why U.S. Manufacturing Firms Patent or Not*, NBER Working Paper, Vol. 7552 (National Bureau of Economic Research, Cambridge 2000)
44.10 K. Blind: A taxonomy of standards in the service sector, Serv. Ind. J. **26**(4), 397–420 (2006)
44.11 P. Wakke, K. Blind, H. De Vries: Driving factors for service providers to participate in standardization: Insights from the Netherlands, Ind. Innov. **22**(4), 299–320 (2015)
44.12 I. Heras-Saizarbitoria, O. Boiral: ISO 9001 and ISO 14001: Towards a research agenda on management system standards, Int. J. Manag. Rev. **15**(1), 47–65 (2013)
44.13 J. Llach, F. Marimon, M. Bernardo: ISO 9001 diffusion analysis according to activity sectors, Ind. Manag. Data Syst. **111**(1/2), 298–316 (2011)
44.14 F. Marimon, J. Llach, M. Bernardo: Comparative analysis of diffusion of the ISO 14001 standard by sector of activity, J. Clean. Prod. **19**(15), 1734–1744 (2011)
44.15 R. Hahn: ISO 26000 and the standardization of strategic management processes for sustainability and social responsibility, Bus. Strategy Environ. **22**, 442–455 (2013)
44.16 J. West: Seeking open infrastructure: Contrasting open standards, Open Source Open Innov. **12**, 1–16 (2007)
44.17 K. Blind, S. Gauch: Trends in ICT standards in European standardisation bodies and standards consortia, Telecommun. Policy **32**(7), 503–513 (2008)
44.18 P.M. Wiegmann, H.J. De Vries, K. Blind: Multi-mode standardisation: A critical review and a research agenda, Res. Policy **46**, 1370–1386 (2017)
44.19 K. Blind, A. Mangelsdorf: Motives to standardize: Empirical evidence from Germany, Technovation **48/49**, 13–24 (2016)
44.20 J. Baron, T. Pohlmann, K. Blind: Essential patents and standard dynamics, Res. Policy **45**, 1762–1773 (2016)
44.21 OECD: *Sectoral Studies*, The OECD Report on Regulatory Reform, Vol. I (OECD, Paris 1997)
44.22 CEN: *Standardization and Related Activities—General Vocabulary EN 45020:2006* (European Committee for Standardization, Brussels 2006)
44.23 G.M.P. Swann: *The Economics of Standardization* (Department of Trade and Industry, London 2000)
44.24 K. Blind: *The Economics of Standards: Theory, Evidence and Policy* (Edward Elgar, Cheltenham 2004)
44.25 OECD/Eurostat: *Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation*, 4th edn. (OECD/Eurostat, Paris/Luxembourg 2018), https://doi.org/10.1787/9789264304604-en
44.26 SNV Schweizerische Normen-Vereinigung: The basics of standardization, https://www.snv.ch/fileadmin/snv/Normung/Dokumente/1x1_der_Normung_-_alle_Sprachen/120509_SNV_WEB_1x1_Brosch_E.pdf (2012)
44.27 K. Blind: Standardisation: A catalyst for innovation, inaugural address given at Rotterdam School of Management, Erasmus University (2009)
44.28 K. Blind, S. Gauch: Research and standardisation in nanotechnology: Evidence from Germany, J. Technol. Transf. **34**(3), 320–342 (2009)
44.29 A.-M. Großmann, E. Filipović, L. Lazina: The strategic use of patents and standards for new product development knowledge transfer, R&D Management **46**, 312–325 (2016)
44.30 P.A. David, S. Greenstein: The economics of compatibility standards: An introduction to recent research, Econ. Innov. New Technol. **1**(1/2), 3–41 (1990)
44.31 A. Zi, K. Blind: Researchers' participation in standardisation: A case study from a public research organisation in Germany, J. Technol. Transf. **40**(2), 346–360 (2015)
44.32 M. Raven, K. Blind: The characteristics and impacts of scientific publications in biotechnology research referenced in standards, Technol. Forecast. Soc. Change **115**, 167–179 (2017)
44.33 M. Rysman, T.S. Simcoe: Patents and the performance of voluntary standard-setting organizations, Manag. Sci. **54**(11), 1920–1934 (2008)
44.34 T. Pohlmann, K. Blind: *Landscaping Study on Standard Essential Patents (SEPs), Study on behalf of DG GROW* (iPlytics, Berlin 2017)
44.35 R. Bekkers, R. Bongard, A. Nuvolari: An empirical study on the determinants of essential patent claims in compatibility standards, Res. Policy **40**(7), 1001–1015 (2011)
44.36 F. Berger, K. Blind, N. Thumm: Filing behaviour regarding essential patents in industry standards, Res. Policy **41**, 216–225 (2012)
44.37 B. Kang, K. Motohashi: Essential intellectual property rights and inventors involvement in standardization, Res. Policy **44**(2), 483–492 (2015)
44.38 B. Kang, R. Bekkers: Just-in-time patents and the development of standards, Res. Policy **44**, 1948–1961 (2015)
44.39 K. Hussinger, F. Schwiebacher: The market value of technology disclosures to standard setting organizations, Ind. Innov. **22**(4), 321–344 (2015)
44.40 T. Pohlmann, P. Neuhäussler, K. Blind: Standard essential patents to boost financial returns, R&D Management **46**(S2), 612–631 (2016)
44.41 K. Blind: Explanatory factors for participation in formal standardisation processes: Empirical evidence at firm level, Econ. Innov. New Technol. **15**(2), 157–170 (2006)
44.42 K. Blind, A. Mangelsdorf: Alliance formation of SMEs: Empirical evidence from standardization committees, IEEE Trans. Eng. Manag. **60**(1), 148–156 (2013)
44.43 A.E. Leiponen: Competing through cooperation: The organization of standard setting in wireless telecommunications, Manag. Sci. **54**(11), 1904–1919 (2008)

44.44 P. Wakke, K. Blind, F. Ramel: The impact of participation within formal standardization on firm performance, J. Prod. Anal. **45**(3), 317–330 (2016)
44.45 K. Blind, J. Pohlisch, A. Zi: Publishing, patenting and standardization: Motives and barriers of scientists, Res. Policy **47**(7), 1185–1197 (2018)
44.46 K. Blind, A. Großmann: An exploratory analysis of company standards–Evidence from the German standardization panel. In: *Proc. EURAS*, ed. by I. Mijantovic, K. Jakobs (2014) pp. 33–47
44.47 H. De Vries: *Standardization–A Business Approach to the Role of National Standardization Organisations* (Kluwer Academic, Dordrecht 1999)
44.48 K. Blind, A. Jungmittag: The impact of standards and patents on macroeconomic growth: A panel approach covering four countries and twelve sectors, J. Product. Anal. **1**, 51–60 (2008)
44.49 G.P. Swann: *International Standards and Trade: A Review of the Empirical Literature*, Working Papers, Vol. 97 (OECD, Paris 2010)
44.50 K. Blind, A. Mangelsdorf, C. Niebel, F. Ramel: Standards in the global value chains of the European single market, Rev. Int. Political Econ. **25**, 28–48 (2018)
44.51 J. Baron, D. Spulber: Technology standards and Standards Organizations: Introduction to the Searle Center Database, J. Econ. Manag. Strat. **27**(3), 462–503 (2015)
44.52 I. Guler, M.F. Guillén, J.M. Macpherson: Global competition, institutions, and the diffusion of organizational practices: The international spread of ISO 9000 quality certificates, Adm. Sci. Q. **47**(2), 207–232 (2002)
44.53 B. Manders, H. De Vries, K. Blind: ISO 9001 and product innovation: A literature review and research framework, Technovation **48/49**, 41–55 (2016)
44.54 S. Lim, A. Prakash: Voluntary regulations and innovation: The case of ISO 14001, Public Adm. Rev. **74**, 233–244 (2014)
44.55 M. Ehrich, A. Mangelsdorf: The role of private standards for manufactured food exports from developing countries, World Dev **101**, 16–27 (2018)
44.56 A. Ploschka: Standards and their application in hydraulic products. In: *Proc. EURAS*, ed. by K.B. Skriver, K. Jakobs, J. Jerlang (2015) pp. 323–340
44.57 G.M.P. Swann, K. Lambert: Why do standards enable and constrain innovation? In: *Proc. EURAS*, ed. by J.-C. Graz, K. Jakobs (2010) pp. 357–376
44.58 M. Frenz, R. Lambert: The economics of accreditation, NCSLI Measure **9**(2), 42–50 (2014)
44.59 C. Rammer, T. Crass, T. Doherr, M. Hud, P. Hünermund, Y. Iferd, C. Köhler, B. Peters, T. Schubert: *Innovationsverhalten der deutschen Wirtschaft: Indikatorenbericht zur Innovationserhebung 2015* (Zentrum für Europäische Wirtschaftsforschung, Mannheim 2016)
44.60 C. Rammer, T. Schubert, P. Hünermund, M. Köhler, Y. Iferd, B. Peters: *Dokumentation zur Innovationserhebung 2015* (Zentrum für Europäische Wirtschaftsforschung, Mannheim 2015)
44.61 K. Blind, J. Pohlisch: The influence of patents and standards as knowledge sources on innovation. In: *Proc. EURAS*, ed. by K. Jakobs, A. Mione, A.-F. Cutting-Decelle, S. Mignon (2016) pp. 69–86
44.62 K. Blind, J.-A. Müller: FNS und TU Berlin stellen die Ergebnisse der vierten Befragungsrunde des Deutschen Normungspanels (DNP) mit Fokus auf Digitalisierung vor, DIN-Mitteilungen **6**, 21–25 (2016)
44.63 Commission of Experts for Research and Innovation: *Report on Research, Innovation and Technological Performance in Germany 2017* (EFI, Berlin 2017) http://www.e-fi.de/fileadmin/Gutachten_2017/EFI_Report_2017.pdf
44.64 International Organisation for Standardisation (ISO): ISO: A global network of national standards bodies, https://www.iso.org/members.html
44.65 International Organization for Standardization (ISO): Annual report 2016: Navigating a world in transition, https://www.iso.org/files/live/sites/isoorg/files/about%20ISO/annual_reports/en/annual_report_2016_en.pdf (2017)
44.66 DIN-Mitteilungen Dezember (2017) p. 122
44.67 International Organization for Standardization (ISO): The ISO survey, http://www.iso.org/iso/iso-survey (2016)
44.68 H. Zoo, H.J. De Vries, H. Lee: Interplay of innovation and standardization: Exploring the relevance in developing countries, Technol. Forecast. Soc. Change **118**(C), 334–348 (2017)
44.69 M.P. Hekkert, R. Suurs, S.O. Negro, K. Kuhlmann, R. Smits: Functions of innovation systems: A new approach for analysing technological change, Technol. Forecast. Soc. Change **74**(4), 413–432 (2007)
44.70 K. Blind: From standards to quality infrastructure: A review of impact studies and an outlook. In: *The Law, Economics and Politics of International Standardisation*, ed. by P. Delimatis (Cambridge Univ. Press, Cambridge 2015) pp. 42–57
44.71 J.L. Guasch, J.L. Racine, I. Sánchez, M. Diop: *Quality Systems and Standards for a Competitive Edge* (World Bank, Washington 2007)
44.72 C. Sanetra, R.M. Marbán: *The Answer to the Global Quality Challenge: A National Quality Infrastructure* (Physikalisch-Technische Bundesanstalt, Braunschweig 2007)
44.73 OECD, ETF, EU, EBRD, SEECEL: *SME Policy Index: Western Balkans and Turkey 2016: Assessing the Implementation of the Small Business Act for Europe* (OECD, Paris 2016)
44.74 K. Blind, A. Mangelsdorf, J. Pohlisch: The effects of cooperation in accreditation on international trade: Empririca evidance on ISO 9000 artifications, J. Product. Econ. **198**, 50–59 (2018)
44.75 J. Gonçalves, J. Peuckert: The contributions of quality infrastructure to national innovation systems: Implications for development policy. In: *Innovation Systems and Capabilities in Developing Regions*, ed. by W. Siyanbola (Gower, Burlington 2012) pp. 35–48
44.76 K. Blind: The impact of standardisation and standards on innovation. In: *Handbook of Innovation*

Policy Impact, ed. by J. Edler, P. Cunningham, A. Gök, P. Shapira (Edward Elgar, Cheltenham 2016) pp. 423–449

44.77 K. Blind: The economic functions of standards in the innovation process. In: *Handbook of Innovation and Standards*, ed. by R. Hawkins, K. Blind, R. Page (Edward Elgar, Cheltenham 2017) pp. 38–62

44.78 K. Blind, S. Petersen, C. Riillo: The impact of standards and regulation on innovation in uncertain markets, Res. Policy **46**(1), 249–264 (2017)

44.79 M. Perkmann, V. Tartari, M. McKelvey, E. Autio, A. Broström, P. D'Este, R. Fini, A. Geuna, R. Grimaldi, A. Hughes, S. Krabel, M. Kitson, P. Llerena, F. Lissoni, A. Salter, M. Sobrero: Academic engagement and commercialisation: A review of the literature on university–industry relations, Res. Policy **42**(2), 423–442 (2013)

44.80 J. West, M. Bogers: Leveraging external sources of innovation: A review of research on open innovation, J. Prod. Innov. Manag. **31**(4), 814–831 (2014)

44.81 M. Bogers, A.-K. Zobel, A. Afuah, E. Almirall, S. Brunswicker, L. Dahlander, L. Frederiksen, A. Gawer, M. Gruber, S. Haefliger, J. Hagedoorn, D. Hilgers, K. Laursen, M.G. Magnusson, A. Majchrzak, I.P. McCarthy, K.M. Moeslein, S. Nambisan, F.T. Piller, A. Radziwon, C. Rossi-Lamastra, J. Sims, A.L.J. Ter Wal: The open innovation research landscape: Established perspectives and emerging themes across different levels of analysis, Ind. Innov. **24**(1), 8–40 (2017)

44.82 European Commission: *A Lead Market Initiative for Europe*, *COM (2007) 860 final* (European Commission, Brussels 2007)

44.83 J. Edler, L. Georghiou, E. Uyarra, K. Blind: Evaluating the demand side: New challenges for evaluation, Res. Eval. **21**, 33–47 (2012)

44.84 CESS/Oxford Research: *Final Evaluation of the Lead Market Initiative*, *Final Report* (Publications Office of the European Union, Luxembourg 2011)

44.85 European Commission: *Horizon 2020–The Framework Programme for Research and Innovation*, *COM (2011) 808 final* (European Commission, Brussels 2011)

44.86 European Commission: *Proposal for a Regulation of the European Parliaments and of the Council Establishing Horizon 2020–The Framework Programme for Research and Innovation (2014–2020)*, *COM (2011) 809 final* (European Commission, Brussels 2011)

44.87 European Commission: *Proposal for a Regulation of the European Parliaments and of the Council Laying Down the Rules for the Participation and Dissemination in Horizon 2020–The Framework Programme for Research and Innovation (2014–2020)*, *COM (2011) 810 final* (European Commission, Brussels 2011)

44.88 M.G. Löhe, K. Blind: Regulation and standardization of data protection in cloud computing. In: *ITU Kaleidosc. Trust Inf. Soc. K-2015*, *Barcelona* (2015) pp. 1–6

44.89 N. Egan, J. Pelkmans: TTIP's hard core: Technical barriers to trade and standards, Paper No. 13 in the CEPS-CTR project 'TTIP in the Balance' and CEPS, Special Report No. 117/August 2015 (2015)

44.90 K. Blind, J.-A. Mueller: The role of standards in the policy debate on the EU-US trade agreement, J. Policy Modeling (2018), https://doi.org/10.1016/j.jpolmod.2018.09.006

44.91 K. Goluchowicz, K. Blind: Identification of future fields of standardisation: An explorative application of the Delphi methodology, Technol. Forecast. Soc. Change **78**, 1526–1541 (2011)

Knut Blind

Institute of Technology and Management
Technische Universität Berlin
Berlin, Germany
knut.blind@tu-berlin.de

Knut Blind is Professor of Innovation Economics at the Technische Universität Berlin. He studied Economics, Political Science and Psychology at Freiburg University. He was awarded a BA at Brock University (Canada). He received his Diploma in Economics and his Doctoral degree from Freiburg University. In April 2006, Knut Blind was appointed Professor of Innovation Economics at the Technische Universität Berlin. He also held an Endowed Chair on Standardisation at Rotterdam School of Management between 2008 and 2016.

Detailed Contents

Part B Advancement of Methodology for Research Assessment

Part D New Indicators for Research Assessment

Part F Patent System, Patents and Economics

Subject Index

A

B

D

E

F

G

H

I

J

K

L

M

N

Q

R

S

T

U

V

W