

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 ... | 687 |
| 26.2 | Research Evaluation: Principles, Frameworks, and Challenges | 688 |
| 26.2.1 | Origins: The Altmetrics Manifesto | 688 |
| 26.2.2 | Standards, Critiques and Guidelines.... | 689 |
| 26.2.3 | Individual-Level Metrics..... | 689 |
| 26.2.4 | Responsible Metrics | 690 |
| 26.3 | Social Media Data and Indicators | 691 |
| 26.3.1 | Social Media Metrics Tools | 691 |
| 26.3.2 | Characterizing Interactions and Users in Social Media Metrics | 693 |
| 26.4 | Conceptualizing Social Media Metrics for Research Evaluation and Management | 694 |
| 26.4.1 | Validity and Reliability of Social Media Metrics..... | 694 |
| 26.4.2 | Homogeneity (or Heterogeneity) of Altmetric Indicators | 695 |
| 26.5 | Data Issues and Dependencies of Social Media Metrics | 696 |
| 26.6 | Conceptualizing Applications of Social Media Metrics for Research Evaluation and Management | 696 |
| 26.6.1 | Descriptive Social Media Metrics..... | 697 |
| 26.6.2 | Comparative Indicators | 704 |
| 26.7 | Prospects for Social Media Metrics in Research Evaluation | 705 |
| 26.7.1 | Understanding the Nature of Social Media Metrics for Research Evaluation | 705 |
| 26.7.2 | Proposing Alternative Forms of Research Evaluation Based on Social Media Metrics..... | 707 |
| 26.8 | Concluding Remarks | 708 |
| | References | 709 |

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

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 in-

dicators 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

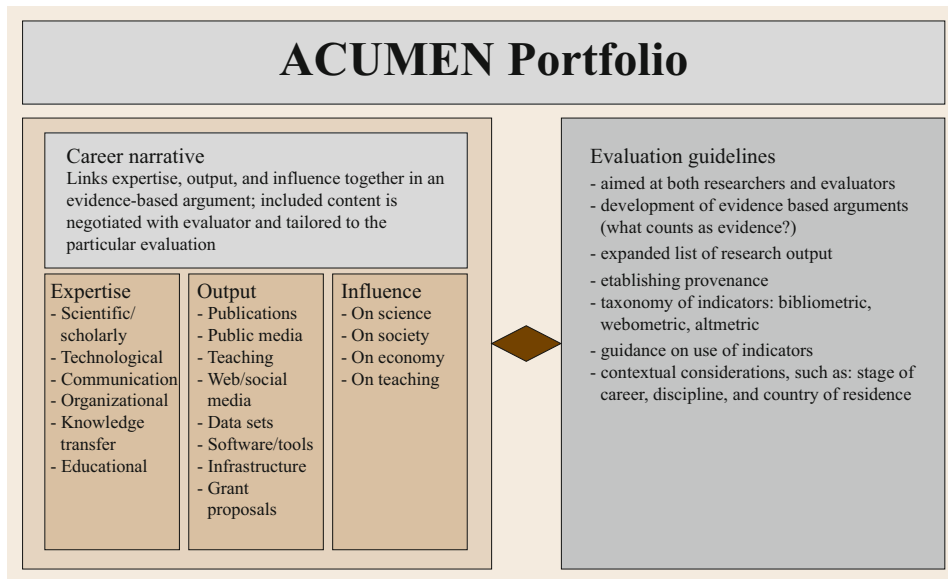


Fig. 26.1 The ACUMEN portfolio [26.20]

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 CiteU-Like (<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.
- *Almetrics data aggregators*. These are tools such as Almetric.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. Almetric.com provides a composite weighted indicator (*Almetric 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 (publi-

cations 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]. Almetric.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 sug-

gested 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 it 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 |
|---|---|
| <ul style="list-style-type: none"> ● Descriptive social media indicators (nonnormalized), e. g.: <ul style="list-style-type: none"> – Total counts, coverage – Trend analyses ● Social media metrics landscapes <ul style="list-style-type: none"> – Thematic landscapes – Geographic landscapes ● Network approaches, e. g., communities of attention, Twitter coupling, hashtag coupling | <ul style="list-style-type: none"> ● Normalized indicators, e. g.: <ul style="list-style-type: none"> – Mendeley field-normalized indicators – Percentile-based indicators (e. g., Altmetric Attention Score) ● Social media-based <i>factors</i> (e. g., Twiimpact 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 | <i>P</i> | <i>P</i> (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

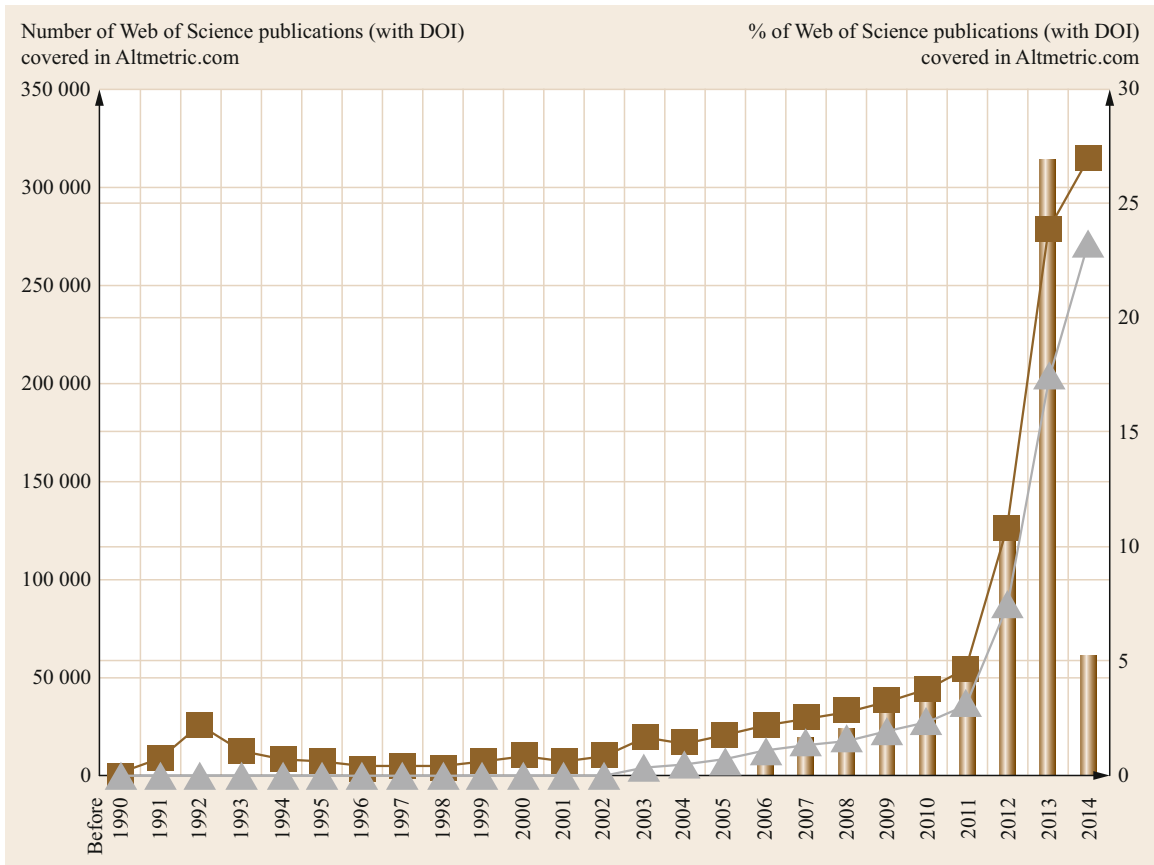


Fig. 26.2 Number and share of publications from Web of Science (DOI) with coverage in Altmetric.com, 1980–2013 (after [26.60], with permission from Wiley). Altmetric.com started their data collection in July 2011

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-

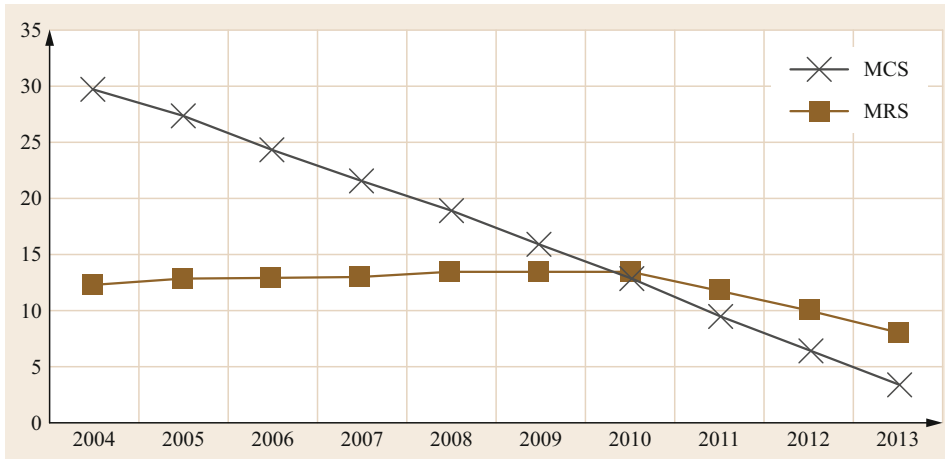


Fig. 26.3 Distributions of mean readership score (MRS) and mean citation score (MCS) indicators for the WoS publications over time (x axis shows the publication years and y axis shows the mean scores for citations and readership) (after [26.57], with permission from Wiley)

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 bibliometric 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].

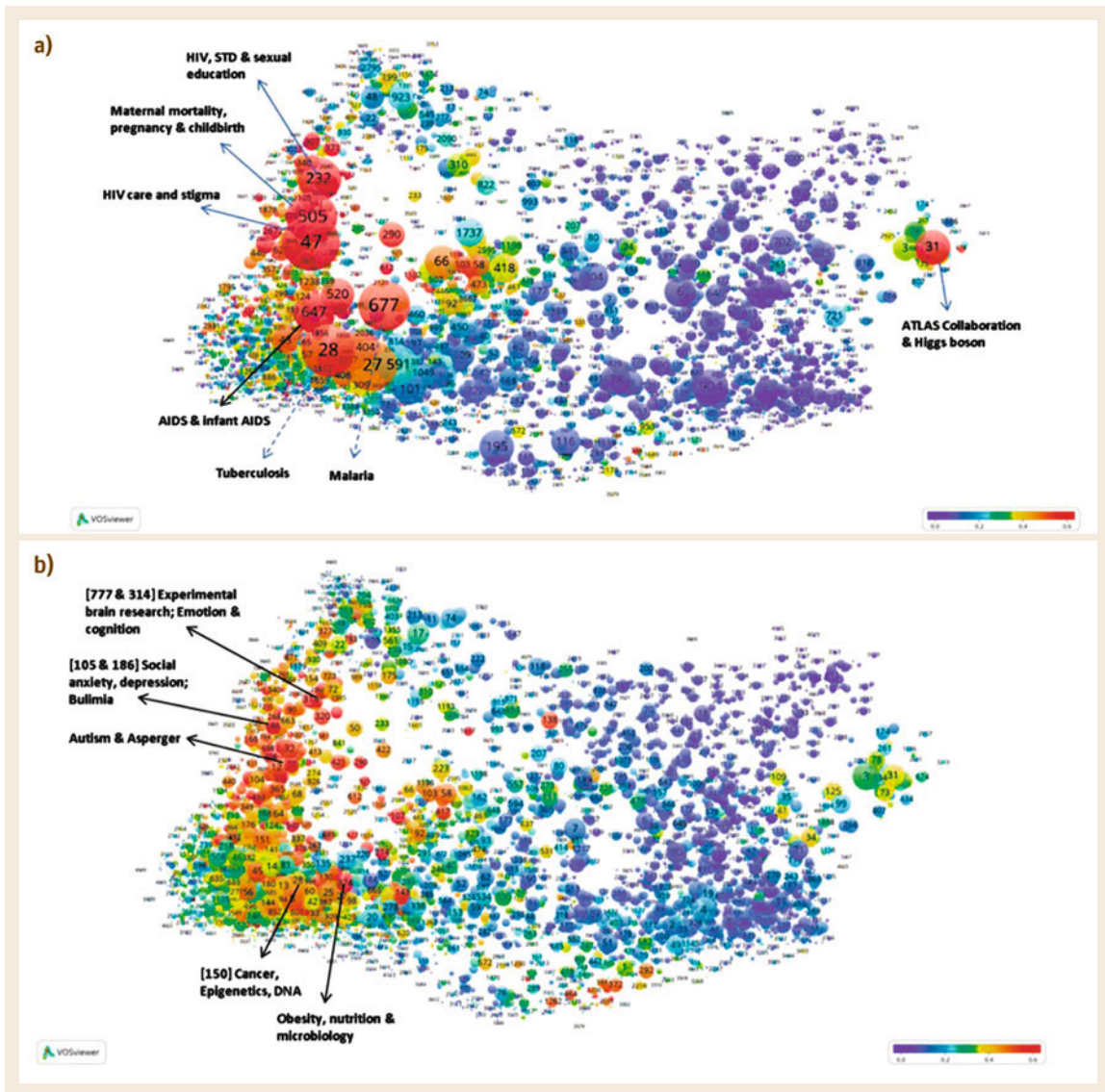


Fig. 26.4a,b Tweets thematic landscape of African publications (a) and EU28 publications (b). Nodes represent fields (clusters of publications closely related by direct citation relations) and position on the map by the strength of their citation relations

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 land-

scapes. 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 possibili-

ties 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

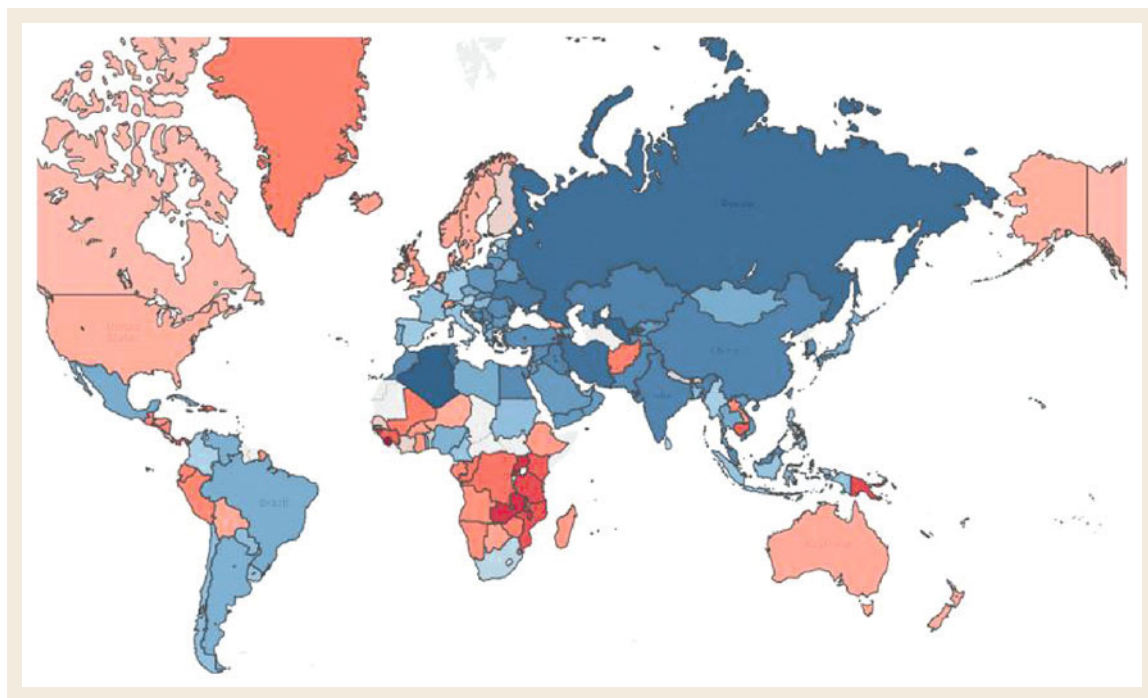


Fig. 26.5 Global map of the share of WoS publications (with a DOI/PMID, period 2012–2014) with at least one Twitter mention ($PP(tw1)$) across the countries of the authors. Threshold for red/blue differences is 34% (i. e., $PP(tw1) < 34\%$ blue, $PP(tw1) \geq 34\%$ red)

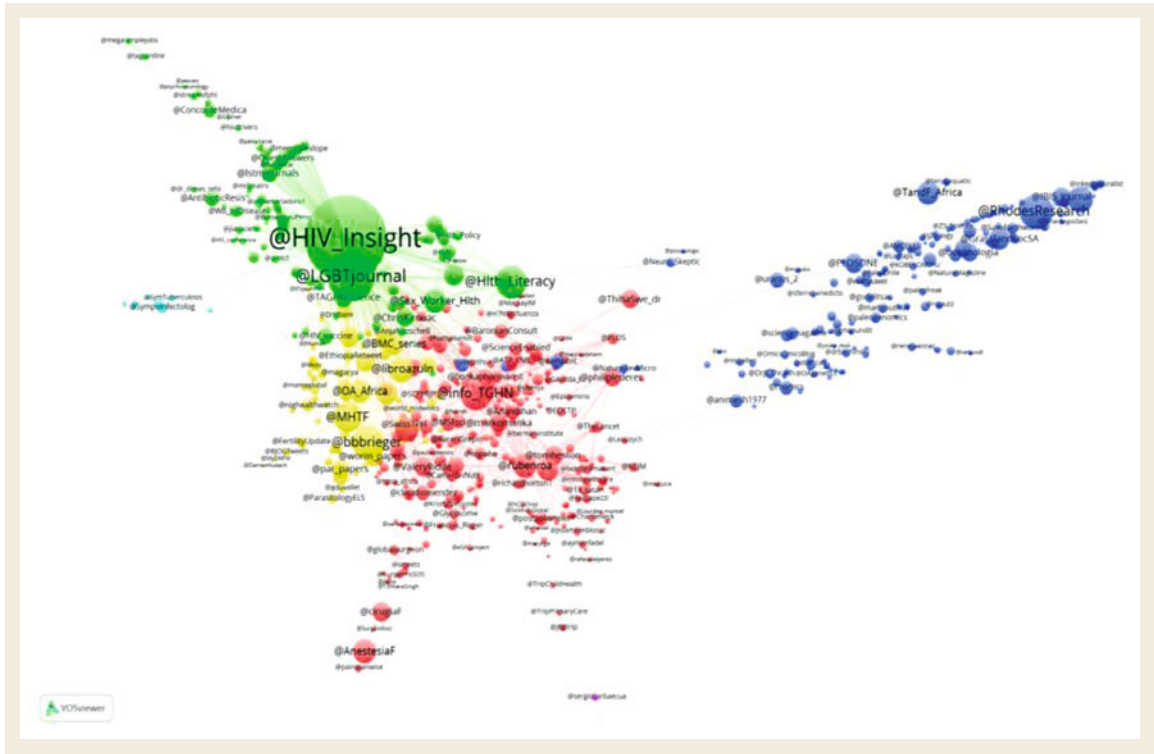


Fig. 26.6 Main Twitter community of attention map of African publications. Nodes are Twitter users; linkages/proximity of the nodes is determined by the number of common publications they have tweeted. Position of nodes in the map: VOSviewer clustering method

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 linking 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 co-citation [26.88, 89]. Those publications with high co-occurrence 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

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 (Twimfact 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 schol-

arly 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.

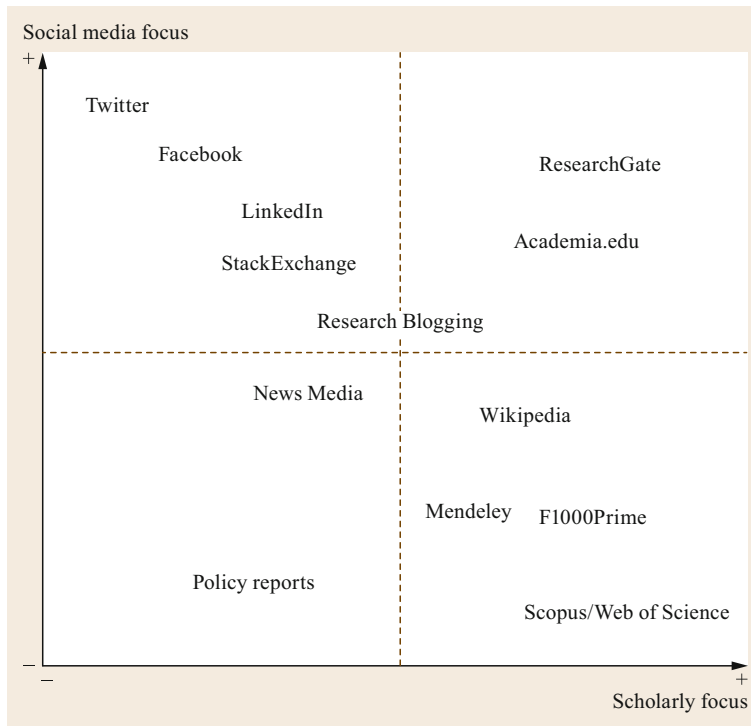


Fig. 26.8 Metrics characterized by their focus: social media or scholarly

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 re-

search 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

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

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.

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 Pol-

icy 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 ci-

- tations?, *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 post-docs? 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 multi-disciplinary 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* **4**1, 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, Bloomington* (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 Analysis 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. Mafrahi, 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.