Chapter 6
Dynamic Scientific Co-Authorship Networks

Franc Mali, Luka Kronegger, Patrick Doreian, and Anuška Ferligoj

6.1   Introduction

Network studies of science greatly advance our understanding of both the knowledge-creation process and the flow of knowledge in society. As noted in the introductory chapter, science can be defined fruitfully as a social network of scientists together with the cognitive network of knowledge items (Börner et al. 2012). The cognitive structure of science consists of relationships between scientific ideas, and the social structure of science is mostly manifested as relationships between scientists. Here, we confine our attention to these relations. In particular, co-authorship networks among scientists are a particularly important part of the collaborative social structure of science. Modern science increasingly involves "collaborative research", and this is integral to the social structure of science. Ziman argues that the organizational units of modern science are groups and not individuals (Ziman 1994, p. 227).[1] Namely, co-authorship in science presents a

---

[1]Co-authorship in science is not the only form of scientific collaboration. de Haan (1997) suggests six operationalized indicators of collaboration between scientists: co-authorship; shared editorship of publications; shared supervision of PhD projects; writing research proposal together; participation in formal research programs; and shared organization of scientific conferences. As this list suggests, there are many cases of scientific collaborations that do not result in co-authored publications (Katz and Martin 1997; Melin and Persson 1996; Laudel 2002). Laudel (2002) reports that about half of scientific collaborations are invisible in formal communication channels either

F. Mali (✉) · L. Kronegger (✉) · A. Ferligoj
Faculty of Social Sciences, University of Ljubljana, Kardeljeva pl. 5, 1000 Ljubljana, Slovenia
e-mail: franc.mali@fdv.uni-lj.si; luka.kronegger@fdv.uni-lj.si; Anuska.ferligoj@fdv.uni-lj.si

P. Doreian
Department of Sociology, University of Pittsburgh, 2602 WWPH, Pittsburgh, PA 15260, USA

Faculty of Social Sciences, Kardeljeva pl. 5, 1000 Ljubljana, Slovenia
e-mail: pitpat@pitt.edu

more substantial indicator than just scientific communication in one way or another. In continuation, we focus on the dynamics of different kinds of co-authorship networks.

Over the last 50 years, the study of the dynamics of co-authorship networks has been conditioned by the development of quantitative methodological approaches in various forms that include relatively simple descriptive statistics presented in time-series form, deterministic approaches, and stochastic agent-based modeling of network dynamics. We provide a brief overview of these approaches in this chapter. Many studies of co-authorship networks are typically described and understood in terms of very large networks involving tens of thousands of nodes. Science can be understood as social phenomena involving large numbers of scientists regularly performing specific actions that are consciously coordinated into large schemes (Ziman 2000, p. 4). Different disciplinary approaches allow the use of different statistical quantities to explain the topology of scientific networks. Some of the statistical quantities typically used to describe these networks are purely local. The other statistical quantities correspond to global descriptions. For example, the local property of a unit in the network is vertex degree, defined as the number of ties relating this unit to other units in the network. Corresponding global descriptions of the degree distribution, which is known to have a long tail for a wide range of different networks, can be constructed (see, for example, Lambiotte and Panzarasa 2009.).

Although co-authorship networks may provide a window on patterns of collaboration within science, they have received far less attention than have citation networks in bibliometrics (Newman 2004, p. 5200). There is a basic difference between co-authorship networks and citation networks. Citation networks are not personal social networks, even though they are, in part, the product of social network phenomena involving scientists. They do not capture the social interaction structure usually described in works on co-authorship networks. These social interaction structures are best described by co-authorship networks. The vertices of co-authorship networks represent authors, and two authors are connected by a tie if they co-authored one or more publications. These ties are necessarily symmetric. In citation networks, the vertices represent scientific productions,[2] and the links between them are directed citation ties from one scientific document to other such documents. In that sense, co-authorship networks contain much important information about cooperation patterns among authors as well as the status and locations of authors in the broader scientific community structures. The study of community structures through scientific co-authorship is particularly important

---

because they do not result in co-authored publications or in formal acknowledgments in scientific texts. In this chapter, we will use the term *collaboration* primarily to designate research that results in co-authored publications and other publicly available documents.

[2]We include papers, monographs, short articles, conference presentations, databases and patents within the term 'scientific production.'

Table 6.1 List of major questions and models presented in this section

	Major issues addressed	Key answers/insights
Barabási and Albert (1999)	Ways of modeling cumulative advantage principle in co-authorship networks.	Using the preferential attachment model where a scale-free power-law distribution of the number of co-authors is a consequence of two generic mechanisms: (i) networks expand continuously by the addition of new vertices, and (ii) new vertices attach preferentially to sites that are already well connected.
Watts and Strogatz (1998)	Ways of modeling the clustered structure of co-authorship networks at the macro level.	Small-world model overcomes the gap in clustering of real-world networks in comparison to random networks. Such constructed networks have small average shortest paths and incorporate clusterings (small dense parts of the network) which emerge in social networks.
Lorrain and White (1971), Doreian et al. (2005)	Ways of clustering the units in co-authorship networks regarding the structure of collaboration and representing the obtained clusters with their connections at the macro level.	The procedural goal of blockmodeling is to identify, in a given network, clusters (classes) of units (actors) that share structural characteristics defined in terms of some relation. Each such cluster forms a position. The units within a cluster have the same or similar connection patterns.
Snijders (1996), Snijders et al. (2010)	Ways of modeling the effects of actor characteristics and network positions on network evolution. Ways of modeling network dynamics and testing results using the inferential methods.	Stochastic actor-based modeling for network dynamics is based on longitudinally observed network data. It is meant to represent and model co-evolution of longitudinal network data and actor attributes, and evaluate the results within the framework of statistical inference.

because scientific (sub)disciplines might often display local properties that differ greatly from the properties of the scientific network as a whole(Table 6.1).

This chapter is structured in the following way. Given that we treat co-authorship networks as social networks, we continue this introduction with a definition of a network. In the next section, we offer a brief historical overview of social network analysis with a focus on the dynamics of social networks. Section 6.3 contains an organizing typology of both the content and units of analysis for the topics we consider. Section 6.4 is the core of the chapter and provides an overview of known methodological approaches for studying dynamic scientific co-authorship networks. In the final section, we outline some benefits and limitations of each approach and finish with a statement of some open problems.

6.1.1 Networks as Graphs

A network consists of observed units and the relationships among them. Units can be represented as vertices and relationships (ties) as links. When using this skeleton representation, each network is a graph.

But this is a simplification of a network. Units (vertices) in the network can have properties. There can be multiple types of vertices in the network. An example is a social network where the vertices represent people and the groups to which they belong. Units also have many different properties (e.g. gender, age, income).

The links in networks can also be of different relational types and, further, strength of relationships can be indicated by adding weights. The vertices and links of networks studied in time have additional properties when time is considered. The timing of relational formations and dissolutions can be recorded and modeled. Duration of relational ties becomes another important property of relations when they are present.

The information of a graph can also be presented in a matrix form. The most common presentation is with the adjacency matrix in which there is a row and a column for each vertex. Non-zero entries in the matrix are present when links between two corresponding vertices exist.

Adjacency matrices can be extended further if we want to present more complex graphs. For example, if we want to present a graph with multiple links between the vertices, we associate the entry of a single cell a_{ij} in the adjacency matrix with the number of links between the vertices i and j. For the representation of valued graphs, which are graphs with values on the links, the value of a single cell a_{ij} in the adjacency matrix corresponds to the value on the link between vertices i and j.

6.2 A Brief History of Social Network Analysis

Histories of most entities usually have starting dates. However, establishing a starting date for an academic field is difficult because the contributing strands of ideas and methods for a field begin in different times and different places. Modern social network analysis (SNA) started when four distinct features were explicitly brought together (Freeman 2004). These features are: (i) a focus on structural matters by looking at actors embedded within a set of social relations and ties; (ii) the extensive use of systematic empirical data; (iii) heavy use of graphical imagery; and (iv) having foundations in formal, mathematical, and computational models. Recognizing the combination of these elements as defining social network analysis renders the establishment of a precise date of origin less than important. But, based on Freeman's narrative, a start date in the 1930s for what was to become SNA seems reasonable. What matters far more for the field are the operational ways in which the four core components are combined to help us understand network structures and processes.

Academic fields also require some social organization to support them in order to provide an accepted arena for the exchange of ideas and the development of an identity that nurtures a discipline. These were created for SNA within a span of 4 years. Barry Wellman established the International Network of Social Networks Analysts (INSNA) in 1976. He founded *Connections* a year later as a newsletter to distribute news, ideas, and information to members of the field. Lin Freeman established the flagship journal, *Social Networks*, in 1978. Finally, Russ Bernard and Alvin Wolfe started the annual Sunbelt Social Network Conference in 1980. All four entities have grown in size and influence since they were established. The European Network Conference was started in 1989, and in 1995 the two conferences were combined to form the Annual Sunbelt International Social Network Conference.

If we allow that SNA is what social network analysts do, it does not follow automatically that the field is coherent. Hummon and Carley (1993) examined all of the papers in the first 12 volumes of *Social Networks* to assess the state of the field and established that SNA was an integrated scientific community with a shared paradigm. They used 'main-path analysis,' a technique pioneered by Hummon and Doreian (1989, 1990) that helps study the citation patterns of a field. Hummon and Carley (1993) identified 6 main paths in the literature: (i) Role analysis and blockmodeling; (ii) Methods for network analysis; (iii) Concern with network data; (iv) Biased networks; (v) Attention to structure; and (vi) Analyses of personal networks. Of course, these paths for the movement of SNA intellectual ideas through the literature are linked. Hummon and Carley (1993) noted other features of the field. One was the heavy use of formal, mathematical, and quantitative methods. Another was the creation of substantive network ideas, and a third was the presence of prominent collaborative groups of social network analysts. All are consistent with the practice of 'normal science' in the sense of Kuhn (1996).

On looking at that list of main paths as intellectual foci for SNA, one feature leaps out by its absence: There is little about temporal issues[3] even though main path analysis is an explicitly temporal approach. Up until the beginning of the 1990s, SNA appeared to have had a profoundly static bias. The field's concern was centered primarily – but not exclusively – on social structure and patterns of social structures. Given this, four event streams that can be dated as starting in the 1990s have changed the field dramatically.

The first was a series of three special issues of the *Journal of Mathematical Sociology* (JMS) that appeared in 1996, 2001, and 2003. All three issues, edited by Frans Stokman and Patrick Doreian, were devoted to "network evolution." Based on the intuition that "network processes are series of events that create, sustain and dissolve social structures" (Doreian and Stokman 1997, p. 3), the three special issues had a series of papers that looked at network dynamics and network evolution using a variety of different formal models, simulation methods and statistical models.[4]

---

[3]This is consistent with the observations of Powell et al. (2005).

[4]Volume 30(1) of *Social Networks* (2010) was a special issue devoted to network dynamics that noted the importance of the three JMS special issues with papers building upon some of the earlier work.

The second event was the take-off of exponential random graph models (ergms) for the study of change in social networks. The origins of these models date from an earlier time, including the work of Holland and Leinhardt (1981) and Frank and Strauss (1986). One strand of this line work is founded on Wasserman and Pattison (1996) and Pattison and Wasserman (1999) and takes the form of p*-models. This forms the core of the software called Pnet (Wang et al. 2009), used for estimating ergms. Another strand features the work of Snijders (2001) and takes the form of SIENA (Snijders et al. 2010), also used for estimating ergms for studying the co-evolution of social actors and social networks. Yet another strand of related work is present in Statnet (Handcock et al. 2003) that includes the estimation of ergms. There has been a lively debate and an extensive cross-fertilization and collaboration between the groups centered at the University of Melbourne, the University of Groningen, Oxford University, and the University of Washington regarding ergms.

The third event is the movement of physicists into the realm of social networks, which also started in the 1990s. Bonacich (2004) labeled this as "the invasion of the physicists "in his review of Watts (2003) and Barabási (2002). To the extent that the physicists are inattentive to the substantive content of the SNA and reinvent old – and/or even square – wheels, this is an invasion. However, they also bring with them a variety of new modeling strategies and additional conceptualizations of network phenomena that include 'small-world' networks and 'preferential attachment,' two terms that have made fruitful entrances into SNA. The physicists have focused attention primarily on large networks with a view to delineating and understanding network topologies and dynamics.

The final event started in the early 1990s and resulted in the establishment of generalized blockmodeling (Doreian et al. 2005) as both a generalization and an extension of traditional blockmodeling, the main path in the SNA literature through 1992 identified by Hummon and Carley (1993). Thus far, this approach has been deterministic and not that attuned to network dynamics. Designed to delineate network structures through the use of an expanding collection of block types and types of blockmodels, it has the potential to contribute to the temporal delineation of fundamental network structures.

At face value, the four 'events' and the lines of active research that have followed them are different and could even be viewed as potential rivals. However, it will be unfortunate if they are seen in this fashion. Some of the ideas of physicists can be used to conceptualize mechanisms that can be incorporated into ergms to test these ideas with social network data. It is clear that the efforts of physicists to identify communities in networks have the same intent as blockmodeling. The work of Handcock et al. (2007) on discerning network structure through model-based clustering is also related, in intent, to blockmodeling, and it seems reasonable to couple, in some way, ergms and blockmodels. All of these four strands of research for understanding networks have been mobilized extensively since their first appearance. They have all emerged since Hummon and Carley's (1993) assessment and have the potential to be combined fruitfully in future research. While

these streams of research are changing SNA to focus on network dynamics and network evolution, they do so while embodying all of the four defining features of SNA identified by Freeman (2004).

6.3 Levels of Analysis of Scientific Collaboration

6.3.1 Introduction

Understanding science as a social system implies considering science as fundamentally relational, and as a community-based social activity. "The collegian circles around a scientist refer to those local and distant peers or professional colleagues" (Schott 1993, p. 201). These collegian circles have several properties that vary from one scientist to another. Within social studies of science, there has been a strong interest in the spatial range of the collegian circle with attention given to local, national, or transnational scientific communities. These professional collegian circles in science have several other characteristics that are analytically distinct but, in reality, may be intertwined. Co-authorship networks in science have a "modular structure" (Lambiotte and Panzarasa 2009, p. 181). Understanding this modular structure of scientific networks is especially important because it helps account for the progress of science and the organization of scientific production within disciplinary frameworks. In reality, science never operates as a single community with hundreds of thousands of individual scientists. It is organized by many different networks that cut across the formal boundaries dividing science with regard to disciplinary, sectoral, and geographical levels. Of course, the membership of various networks overlaps considerably. These research networks are also in continuous processes of growth, decline, and dissolution (see, for example, Ziman 2000, p. 46 or Mulkay 1975, p. 519).

Classification of co-authorship networks can be done in several ways. Rogers et al. (2001) suggested a typology based on three features: (1) according to the units of the analysis, including individuals, teams of researchers, and R&D organizations; (2) according to the type of information used to develop the links between units – these might be based on interactions or information sharing or they could be based on positions of people in the social hierarchy; and (3) according to the institutionalized domains to which the authors belong, with an emphasis on intra-organizational or inter-organizational links between them. Sonnenwald (2007) suggested a more general classification to categorize various types of co-authorship networks: between researchers in university and industry sectors, between researchers in various scientific disciplines, and between researchers of various countries. In this section, we prefer to use another categorization, one adapting a suggestion by Andrade et al. (2009) who focused on three dimensions of co-authorship networks with their associated sub-dimensions of intra- and inter-dimensional co-authorship collaboration. The suggested dimensions are:

Table 6.2 Classification of levels of analysis of scientific collaboration

Dimension of the study		Examples of studies
Main dimension	Sub-dimension	
Cross-Disciplinary	Disciplinarity	Interaction links between Australian research networks (Rigby 2005), (see also: Wray 2002; Glänzel and Schubert 2004; Laband and Tollison 2000; Hornbostel 1997)
	Inter-disciplinarity	Interdisciplinary research analysis in French laboratories (Sigogneau et al. 2005) (see also Gibbons et al. 1994; Etzkowitz and Leydesdorff 2001; Qin et al. 1997; Braun and Schubert 2003)
Cross-Sectoral	Intramural	Academic research networks analysis (Lowrie and McKnight 2004; Wray 2002)
	Extramural	R&D cooperation models between industry and universities in Belgium (Veugelers and Cassiman 2005)
Cross-National	National	The interaction between immunology research institutes in Germany, due to their geographical location (Havemann et al. 2006)
	International	Comparative analysis of several countries of their international/national collaborated publications (Glänzel and Schubert 2005)

disciplinary with sub-dimensions of interdisciplinary and intradisciplinary, sector with intersector and intrasector, and geographic with international and intranational sub-dimensions. These are presented in Table 6.2.

6.3.2 The Cross-Disciplinary Level

For the cross-disciplinary level, given the presence of disciplinarity, there is a basic distinction between collaboration inside discipline (intra-disciplinarity) and collaboration between disciplines (inter-disciplinarity).

6.3.2.1 Disciplinarity

As stated in the introductory chapter of this book (see page xi et sqq.), "an academic discipline, or field of study, is a branch of knowledge which is taught and researched at the college or university level. Disciplines are defined (in part) and recognized by the academic journals in which research is published, and the learned societies and academic departments or faculties to which their practitioners [researchers] belong" (Börner et al. 2012). Many theorists of science have noted that all scientific disciplines are intellectually (cognitive) and socially structured (Fuchs 1992; Whitley 1984). Scientific disciplines represent institutional and

organizational frameworks within which their intellectual products and cognitive styles are connected to the social structures, mode, and organization of the production of that knowledge. One of the basic characteristics of modern academic scientific communities is that they are still sharply differentiated and structured in terms of disciplines. Individual scientific disciplines can thus be seen as distinct intellectual and social organizational contexts.

Although co-authorship publishing is more common in the natural sciences than in the social sciences, it is continuously increasing in all main scientific areas (Wray 2002; Glänzel and Schubert 2004; Laband and Tollison 2000; Hornbostel 1997). Collaboration, operationalized through co-authorship, is now normative behavior and ubiquitous for practically all scientific disciplines (e.g., over 95% of articles in major periodicals in physics, biochemistry, biology and chemistry are co-authored (Braun-Munzinger 2009)).

6.3.2.2 Interdisciplinarity

In the last two decades, interdisciplinary collaboration has increased dramatically (see, for example, Gibbons et al. 1994; Etzkowitz and Leydesdorff 2001). This phenomenon is broadly discussed in Chap. 1 with attention focused on a tendency of modern science to form heterogeneous (interdisciplinary) teams of researchers solving pressing social problems and with higher accountability requirements (Börner et al. 2012). These attempts have been made to bridge narrow disciplinarities in science. An important feature stimulating interdisciplinary collaboration in modern science is the demand for innovations resulting from the juxtaposition of ideas, tools, and scholars from different scientific domains. Today, there is an overall agreement that inter-disciplinary links are vital for scientific progress because they have the potential to bring unprecedented intellectual and technical power. For example, the converging technologies of the NBIC fields (i.e. nanotechnology, biotechnology, information sciences, and cognitive sciences) are an example of new interdisciplinary research from fields that previously showed limited interdisciplinary connections (see, for example, Buter et al. 2010).

We know that different organizational and cognitive problems make the development of interdisciplinary research particularly challenging. Interdisciplinarity requires extensive networks of scientists and concepts, considerable time investments, and a need for researcher mobility between disciplines. As noted by Bordons and her collaborators, while collaboration among scientists from different disciplines is widespread, measuring it is not easy (Bordons et al. 2004, p. 441). Using bibliometrics, measurement of interdisciplinarity in publications can be approached from different perspectives that include co-authored publications among scientists from different disciplines, co-occurrence of several classification codes in publications, the interdisciplinary nature of journals, and the presence of cross-disciplinary references or citations. The most often used bibliometric indicator of such collaboration is the percentage of co-authored interdisciplinary publications. Yet, computing this percentage is affected by many factors, including the nature of

the organization of scientific communities, R&D policy orientations, and the chosen operationalization of concepts (e.g., the classification scheme of disciplines that is used (Qin et al. 1997; Braun and Schubert 2003)).

6.3.3 The Cross-Sectoral Level

There is a basic difference between collaborations inside the academic scientific community (intramural cooperation) and collaborations between academic science, industry, and governmental bodies (extramural collaboration). Intramural networks in science are usually defined by collaboration within one department, research group, or institute. Extramural collaborations, on the other hand, consider also cooperation between different sectors (see, for example, Glänzel and Schubert 2004).

6.3.3.1 Intramural Collaborations (Intra-Sectoral Collaboration)

In modern science, the establishment of intra-mural networks is the result of the increased processes of professionalization of recent scientific activity. This has led to a large change in the organizational structure of science, and it's worth repeating Ziman's insight: "the organizational units of modern science are not individuals but groups" (Ziman 1994, p. 227). The organization of R&D activity in academic scientific institutions has created typical team structures – for example, modern research groups consist of principal investigators, co-principal investigators, junior researchers, post-docs, and doctoral students. Price suggested that research collaboration is, in part, a response to the shortage of scientists, which allows them to become "fractional" scientists (Price and Beaver 1966).

6.3.3.2 Extramural Collaborations (Cross-Sectoral Collaboration)

Cooperation between different sectors – academic science, industry and government – is now understood as the most important type of extra-mural collaboration. The concepts of 'Mode 2' and the 'Triple Helix' have extended the idea of research networking within and across sectoral borders. Both concepts were developed in the theory of science and R&D policy discussions after 1990. It seems that the concept of Mode 2 knowledge production presented in *The New Production of Knowledge* (Gibbons et al. 1994) became, in the mid-90s, the symbolic banner of new viewpoints regarding scientific collaborations across sectors. The authors of the new (Mode 2) production of knowledge linked the classical concept of transdisciplinarity, defined by common axioms that transcend the narrow scope of disciplinary worldviews through an overarching synthesis, with two additional factors: problem-driven research and research in applied contexts. Similarly, the concept of the Triple Helix has been developed as a neoinstitutional and neoevolutionary model for studying the networks across academic science, industry science, industry, and

government sectors. In these networks, more important than the presence of the agents is the quality of their relationships in a given configuration (Etzkowitz and Leydesdorff 2001). Although there exists already an extensive expert literature on this type of cross-sectoral networks, there is still a lack of bibliometric studies dealing with co-authorship publications between academic and business-enterprise sectors (Lowrie and McKnight 2004, p. 436).

6.3.4 Cross-National Level

Networks of international collaboration have undergone dramatic structural changes in the last few decades. This is in contrast to intranational networks, where the intensity of collaborations have decreased (see, for example: Hoekman et al. 2010; Glänzel and Schubert 2004; Katz 1994).

6.3.4.1 National Collaborations

National collaboration, while visible in domestic contexts, is often regarded as less visible and treated as less important than international collaborations. Often, the observed (relative) high visibility and high citation attractiveness of internationally co-authored publications result in a kind of operational rule: international co-publications appear in high-impact journals and receive more citations than national papers (Glänzel and Schubert 2004). However, the overall visibility and international relevance sometimes does not necessarily reflect the impact of specific papers in solving specific problems at the local level. The results of national collaborations are often incorporated into publications dealing with trans-institutional and international co-authorship (e.g. Munshi and Pant 2004), and are focused directly on collaboration within a specific country (Gossart and Ozman 2009; Mali et al. 2010). Another important aspect of national collaboration results from the international orientation of bibliographic databases like the *Web of Science* or *Scopus*. Often, the results of national co-authorship and the resulting citation patterns, especially for smaller national scientific systems, are less visible in international bibliographic databases. This can be linked to inter-sectoral collaboration within nations. National collaborations across sectors have an additional complexity because they include the involvement of different administrative units. As a result, such research projects are complex and involve a wide range of different outputs of scientific production. Such complex information can only be reported qualitatively or measured through *local* information systems and electronic bibliographic systems; the Slovenian *COBISS*[5] and *SICRIS* databases[6] or the Turkish *ULAKBIM* database.[7]

---

[5]Co-operative Online Biographic System and Services, www.cobiss.si.

[6]Slovenian Current Research Information System, sicris.izum.si

[7]TURKISH ACADEMIC NETWORK and INFORMATION CENTER, www.ulakbim.gov.tr/eng/.

6.3.4.2 International Collaborations

In thinking about the spatial range of collaboration, there is an important difference between geographic distance and crossing international boundaries. While geographical distances between collaborative units in large nations can be long, the geographical distances between collaborating units in different countries can be short. Of the two, crossing international boundaries is more consequential than geographical distance with regard to scientific collaboration. While international scientific collaborations are important generally, they are especially important for small scientific communities such as, for example, the Slovenian scientific community. Isolated and parochial scientific communities are no longer a suitable environment for recognized scientific excellence. Indeed, it can be argued that they never were important in the history of science. Even in the early days of science, different forms of cooperation between scientists of different nations became important elements in the internationalization of science. Even so, because of the new forms of the globalized connections of science, "the traditional cosmopolitan individualism of science is rapidly being transformed in what might be described as transnational collectivism" (Ziman 1994, p. 218).

This trend of increasing international scientific collaboration through co-authorship is especially strong in recent decades. The number of internationally co-authored articles has risen at a faster rate than traditional 'nationally co-authored' articles (Wagner 2005). As noted in the expanding bibliometric literature, the level of international co-authorship is determined by many factors: the size of the country, 'proximity' between countries, either physical (geographical) proximity or immaterial proximity stemming from cultural affinity in a broad (historical, linguistic) sense, socioeconomic factors, changes in electronic forms of communication, and last but not least, the dynamics created by the self-interest of individual scientists pursuing their own careers.

6.4 Methodological Perspectives

6.4.1 Introduction

The development of methodological approaches for analyzing and modeling temporal scientific co-authorship networks has been founded on developments in graph theory and in SNA. To enable the discussion on temporal analysis of network properties, we describe some of the most relevant basic definitions of network properties that we need for understanding the content of coming sections (extensive explanations of SNA terminology and concepts can be found in Wasserman and Faust (1994)):

- *Degree* The degree of a vertex is defined as the number of ties linking this vertex to other vertices in the network. In lay terms, the degree represents the number of

co-authors for each researcher. As a global measure of the whole network, both the average degree or centralization can be considered.

- Network *density* is the proportion of ties in a network relative to the total number possible (sparse *vs.* dense networks).
- *Path* A path is a sequence of vertices and lines from initial vertex to the terminal vertex where all vertices are different.
- *Path length* This is the number of ties it contains.
- A shortest path or a *geodesic distance* between two vertices u and v, denoted as l_{uv}, is the shortest path length between these two vertices. In co-authorship networks, the distance between two authors who collaborate is 1. As a global network characteristic, the average shortest path is usually considered.
- The *global clustering coefficient* can be viewed as the average probability of a tie between co-authors of a selected author. Technically, it measures the density of triangles in the network and therefore measures the extent of densely connected subgroups of vertices in the network.

Another important factor in the development of the field has been access to data sources on scientific collaboration. Before the development of electronic bibliographic databases and, especially, before the implementation of the scientific citation indexes initiated by Garfield (1955) this was a very difficult and time-consuming task. Some of the most visible electronic databases with academic content are the *Web of Science*, *SCOPUS* and *Google Scholar*. A broader discussion on databases and citation indexes can be found in Chap. 7 of this book.

The study of temporal networks, both with regard to network dynamics and network evolution, gained increasing attention since 1996. As noted in Sect. 2, special issues of the *Journal of Mathematical Sociology* (1996, 2001, 2003) were of value. We distinguish three basic approaches for studying dynamic scientific co-authorship networks: (i) basic analysis of network properties using temporal data (usually in the form of a time-series of snapshots, (ii) deterministic approaches to the analysis of scientific co-authorship networks, and (iii) statistical modeling of network dynamics.

6.4.2   Basic Analyses of Network Properties

One of the first analyses of temporal co-publication was presented by Zuckerman (1967) who studied the patterns of productivity, collaboration and co-authorship among Nobel Laureates. While her analysis was quite narrow, in the sense of focusing on a small elite among scientists, this was due to the limitations of the data available at the time. More than 20 years later, Bayer and Smart (1991) focused on publication patterns of US PhD recipients in chemistry in 1960–1962. They used a longitudinal data set spanning from 1962 to 1985 to follow the careers of these researchers through time. Besides single-authored and multi-authored publications, they also distinguished dual-authorship and proposed a typology of

publication patterns of scientists, including six categories which are highly corre-
lated with co-authorship patterns. Researchers were categorized into groups of: Low
producers, Burnouts, Singletons, Team Leaders, Team Players, Doubletons, and
Rank-and-File types. With the development of electronic bibliographic databases,
simple longitudinal analysis of network characteristics (including average vertex
degrees, clustering coefficients, and density) became a common part of most studies
of temporal co-authorship networks (see Babchuk et al. 1999; Glänzel et al. 1999;
Kronegger et al. 2011a).

6.4.3   Deterministic Analysis of Dynamic
Co-Authorship Networks

Although the time dimension is often included in the analysis of co-authorship
networks, it has been mostly restricted to simple temporal time-series descriptions
of some network characteristics and actor attributes. Such basic analyses can be
found in a wide range of publications since results of practically every method for
social network analysis can be represented in time as a series of snapshots. The
most common goal of these methods is delineating structures within co-authorship
networks and accounting for network properties by using some external parameters.
Efforts of researchers to push the methodology further from simple description of
differences between time snapshots are therefore rare and hard to find.

A fruitful way of delineating structures within co-authorship networks is to use
blockmodeling procedures: Let U be a finite set of units and let the units be related
by a binary relation $R \subseteq U \times U$ that determines a network $\mathbf{N} = (U, R)$. One
of the main procedural goals of social network analysis is to identify, in a given
network, clusters of units that share structural characteristics defined in terms of the
relation R. The units within a cluster have the same or similar connection patterns to
the units of other clusters. Result of clustering $\mathbf{C} = \{C_1, C_2, \ldots C_k\}$ is a *partition* of
units $\mathcal{U}$ and relations R into *blocks* $R(C_i, C_j) = R \cap C_i \times C_j$. Each block is defined
in terms of units belonging to clusters C_i and C_j and consists of all arcs from units
in cluster C_i to units in cluster C_j. If $i = j$, the block $R(C_i, C_i)$ is called a *diagonal*
block.

A *blockmodel* consists of structures obtained by shrinking all units from the same
cluster of the clustering $\mathbf{C}$. For an exact definition of a blockmodel, we must be
precise about which blocks produce an arc in the *reduced graph* and which do not.
The reduced graph can be presented also by a relational matrix, called an *image
matrix*.

The partition is constructed by using structural information contained in R only,
and units in the same cluster are equivalent to each other in terms of R alone. These
units share a common structural position within the network.

Blockmodeling, as a set of empirical procedures, is based on the idea that units
in a network can be grouped according to the extent to which they are equivalent in

terms of some *meaningful* definition of equivalence. In general, different definitions of equivalence usually lead to distinct partitions.

Lorrain and White (1971) provided a definition of *structural equivalence*: Units are equivalent if they are connected to the rest of the network in *identical* ways. From this definition it follows that only four possible ideal blocks can appear (Batagelj et al. 1992b; Doreian et al. 2005)

$$\text{Type 0.} \quad b_{ij} = 0 \qquad\qquad \text{Type 2.} \quad b_{ij} = 1 - \delta_{ij}$$
$$\text{Type 1.} \quad b_{ij} = \delta_{ij} \qquad\qquad \text{Type 3.} \quad b_{ij} = 1$$

where δ_{ij} is the Kronecker delta function and $i, j \in C$. The blocks of types 0 and 1 are called the *null* blocks and the blocks of types 2 and 3 the *complete* blocks. For the nondiagonal blocks $R(C_u, C_v), u \neq v$, only blocks of type 0 and type 3 are admissible.

Attempts to generalize the structural equivalence date back at least to Sailer (1978) and have taken various forms. Integral to all formulations is the idea that units are equivalent if they link in equivalent ways to other units that are also equivalent. Regular equivalence, as defined by White and Reitz (1983), is one such generalization.

As was the case with structural equivalence, regular equivalence implies the existence of ideal blocks. The nature of these ideal blocks follows from the following theorem (Batagelj et al. 1992a): Let $\mathbf{C} = \{C_i\}$ be a partition corresponding to a regular equivalence $\approx$ on the network $\mathbf{N} = (U, R)$. Then each block $R(C_u, C_v)$ is either null or it has the property that there is at least one 1 in each of its rows and in each of its columns. Conversely, if for a given clustering $\mathbf{C}$, each block has this property, then the corresponding equivalence relation is a regular equivalence.

Until now, a definition of equivalence was assumed for the *entire* network and the network was analyzed in terms of the permitted ideal blocks. Doreian et al. (2005) generalized the idea of a blockmodel to one where the blocks can conform to more types beyond the three mentioned above, and one where there is no single a priori definition of 'equivalence' for the entire network.

The problem of establishing a partition of units in a network, in terms of a considered equivalence, is a special case of the clustering problem – such that the criterion function reflects the considered equivalence. Such criterion functions can be constructed to reflect the considered equivalence. They measure the fit of a clustering to an ideal one with perfect relations within each cluster and between clusters, according to the selected type of equivalence.

For the direct clustering approach, where an appropriate criterion function that captures the selected equivalence is constructed, a relocation approach can be used to solve the given blockmodeling problem (Doreian et al. 2005).

Inductive approaches for establishing blockmodels for a set of social relations defined over a set of units were discussed above. Some form of equivalence is specified, and clusterings are sought that are consistent with a specified equivalence. Another view of blockmodeling is *deductive* in the sense of starting with a

blockmodel that is specified in terms of substance prior to an analysis. In this case, given a network, a set of types of ideal blocks, and a family of reduced models, a clustering can be determined which minimizes the criterion function. (For details, see, Batagelj et al. 1998; Doreian et al. 2005). Some prespecified blockmodels are designed as hierarchical models with the positions on paths linked by directed ties in a consistent direction. A core-periphery model is such a model where there is one (or several) core position that is strongly connected internally. Peripheral positions are all connected to core positions but not connected to each other, and they are not internally cohesive. There are variations of the core-periphery model; e.g., in which the periphery is not even connected to the core positions. All described blockmodeling approaches are implemented in the program Pajek (Batagelj and Mrvar 2010).

An example of the multi–core–semi–periphery–periphery structure is presented in Fig. 6.1. This specific structure, found in co-authorship networks, consists of: (i) simple cores comprised of scientists co-authoring with all, or most, colleagues in their core (units R3 to R5 and R13 to R16); (ii) bridging cores composed of researchers who connect two or more other simple cores (units R1 and R2); (iii) a semi-periphery made up of authors who co-author with proportionately fewer others in their position and have no systematic patterns of ties to scientists in other
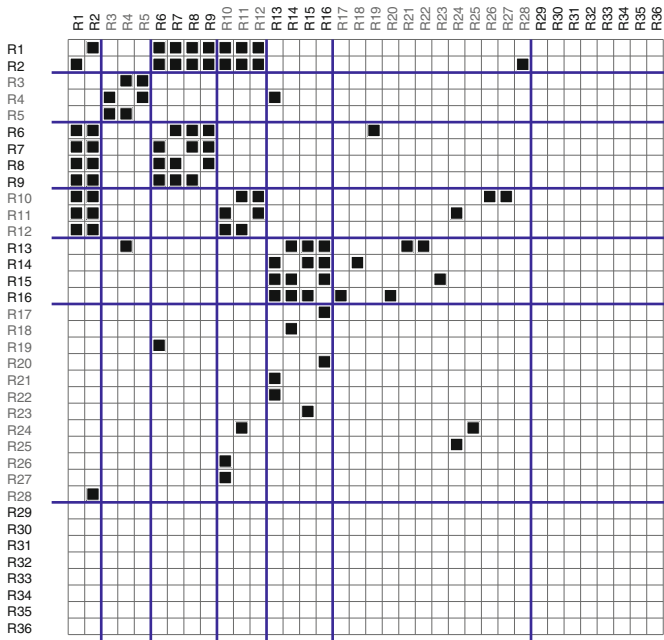
Fig. 6.1 An example of a blockmodel of a network with it's a structure that consists of three parts structure.

positions, and periphery of authors who do not co-author with other researchers from the network.

Several applications of blockmodeling of co-authorship networks have been published in recent years. For example, Said et al. (2008) distinguished several styles of co-authorship, including solo models (no co-authors), mentor models, entrepreneurial models, and team models. They conjectured that certain styles of co-authorship lead to the possibility of group-thinking, reduced creativity, and the possibility of less rigorous reviewing processes. Nooraie et al. (2008) examined co-authorship networks in three Iranian academic research centers in order to find an association between scientific productivity and impact indicators with network features. The collaboration networks within centers shared many structural features, including a "star-like" pattern of relations. Centers with more successful scientific profiles showed denser and more cooperative networks. Kronegger et al. (2011a) distinguished different co-authoring cultures in four scientific disciplines and delineated typical structures of scientific collaboration. They also extended blockmodeling by tracking locations, and hence positions, of authors across different time points.

Another effort to combine a static analysis of complexity at separate time moments with a dynamic analysis was presented by Erten et al. (2004) and by Gansner et al. (2005). They introduced a dynamic extension of multidimensional scaling (Richardson 1938; Torgerson 1952). Multidimensional scaling (MDS) is a set of data analysis techniques designed to display the structure of data in a geometrical picture. The algorithm of dynamic MDS is driven by the minimization of stress measured both within each analyzed year and over consecutive years by optimizing the resulting stress for a three dimensional array. This algorithm was recently implemented in Visone (Leydesdorff and Schank 2008) and used by Leydesdorff (2010) to study co-authorship networks, with additional information on co-word appearance and journal citation indexes. In this paper, he analysed the complete bibliography of Eugene Garfield for the years 1950–2010, graphically presenting its collaboration structure and citation dynamics around Garfields' work mainly dealing with the Science Citation Index.

6.4.4 Modeling Dynamic Scientific Co-Authorship Networks

Here, we present only an overview of modeling temporal co-authorship networks. Static models of macro-level network properties, which are based on stochastic rules of network generation, are discussed first. These have been mainly developed from graph theory by mathematicians and physicists who, with the development of the Internet in 1990, were interested in modeling accessible large real-world networks. The developments led from purely random graphs, built according to the Erdös and Rényi (1959) model, to small-world networks (Watts and Strogatz 1998), and to a range of models based on the concept of preferential attachment (Barabási et al. 2002; Newman 2000).

The idea of finding the rules fostering the growth and development of social networks, or as it was stated, modeling the real world graphs, was widely captured (mostly) by physicists. The basics for any kind of modeling of social networks were provided by the Erdös–Rényi random graph model, which is determined by a number of vertices (n) and the probability (p) that a link exists between two arbitrary vertices. Therefore, each random graph has approximately $p \cdot n(n-1)/2$ undirected links. A single vertex is linked to a binomially distributed number of neighbors. The limiting degree probabilities are Poisson distributed.[8]

The first generalization of the Erdös–Rényi random graph took the form of a *configuration model* where specific degrees are assigned (usually from a pre-specified distribution) to all the vertices which are then randomly linked according to their degree. The construction of the model was proposed by Molloy and Reed (1995) and studied by many authors (see the overview provided in Newman 2003). This solved the problem of degree distribution in real-world graphs usually not having a Poisson distribution, as in the Erdös–Rényis random graph, but not the inability to model the clustered nature of empirical networks.

We consider also a very different approach to modeling social network dynamics, one which returned to and is founded upon ideas within social science. The approach of the physicists has been intent on reproducing the topological form of real-world networks, and it proposes some generic processes of growth and change while ignoring an extensive tradition of sociological and psychological knowledge regarding the behavior of individuals. This alternative (more sociological) approach focuses on single actors and their involvement in the smallest possible social unit of analysis, the dyad. This type of modeling is labeled 'stochastic actor-based modeling' (Snijders 1996). Its purpose is to represent network dynamics on the basis of observed longitudinal data in the form of explicit models and to evaluate them (or a family of models) within the paradigm of statistical inference. This implies that the models are able to represent network change as the result of dynamics being driven by many different tendencies, especially structurally based micro-mechanisms. These mechanisms can be theoretically derived and/or based on empirically established properties in earlier research. Of great importance is that these mechanisms may well operate simultaneously (Snijders et al. 2010). One limitation of these models is that they are restricted to a smaller predetermined number of actors and do not directly consider more global mechanisms of network growth.

6.4.4.1 Modeling "Real-World" Networks

Social studies of science have long had an interest in linking scientific production to the network structures of scientific communities. Different models have been proposed as representations of processes driving co-authorship (as collaboration) in

---

[8]Mathematical notations of models in this section are based on those used by Kejžar (2007).

science that help account for the form of large-scale scientific networks and predict scientific production. One contains an argument that if scientists from particular scientific disciplines (specialties) collaborate with others inside their disciplines, then we would expect to find distinct clusters in the knowledge-production network – exactly the clustering noted in many empirical networks – and this would correspond to small-world network structure (as described below). Alternatively, if the network was generated by preferential attachment (see below) as a mechanism – *where young scientists publish with well-established scientific stars* – then we would expect to find a scale-free network structure whose degree distribution satisfies a power-law. If the network is based on a cross-topic collaboration, then we would not expect to find strong fissures in the network, but instead find a structurally cohesive network (Moody 2004). All of the above-mentioned network structural processes lead to specific dynamics for scientific networks that, in turn, generate distinctive network structures or topologies. These models for generating the structures of large-scale and complex networks can be expected to hold also for co-authorship networks in science. Large-scale co-authorship networks can have local (such as clustering) structural properties as well as global (such as average distance between nodes) structural features. Local and global characteristics of networks help to define network topologies such as "scale-free networks" and "small-world networks." These network topologies are the result of network-generating processes and can lead to further dynamics of these networks in different ways. For example, the principle of preferential attachment to vertices of higher degree leads to a dynamics where "the-rich-get-richer. "In the case of science, this implies that those scientists who experience early success gain higher shares of subsequent rewards. We next consider scale-free and small-world science network structures in more detail.

6.4.4.2 The Small-World Model

The small-world network structure of scientific co-authorship implies network forms where the level of local clustering (one's collaborators are also collaborators with each other) is high, but the average number of steps between clusters is small. In these small-world networks, internal ties to clusters tend to form more cohesive clusters within boundaries, as compared to the more extensive and less cohesive overall networks that include their external ties. According to various social network analysts, the small-world model was inspired by the work of de Sola Pool and Kochen (1978) who partially formalized the much more famous application of Travers and Milgram (1969). It expresses the simple idea that any two individuals, selected randomly from almost anywhere on the planet, are 'connected' via a path of no more than a small number of intermediate acquaintances. The (limited) empirical evidence suggested that this small number is about 6. This notion became a popular idea in the Broadway play named *Six Degrees of Separation*. The first practical evidence for the existence of a small-world phenomenon was first provided by the psychologist Milgram (Berg 2005, p. 46). Milgram's experimental result was

regarded as a good starting point for analyzing the underlying structure of scientific co-authorship.

Later, Watts and Strogatz (1998) formally defined the small-world model in order to construct networks with the following properties that mirror some observed social networks: (i) having short paths between any two vertices (and hence, smaller average lengths for the shortest paths) and (ii) also incorporates clustering (small dense parts of the network). Knowing that geographical proximity of vertices plays a role in the formation of links (especially for humans), they considered a ring-lattice with n vertices. Each vertex had m_{sw} edges to its neighbors. Then they rewired each edge with a probability p_{sw} by relinking the second end of the edge to a randomly chosen vertex. The probability p_{sw} enables this network to vary from an ordered, finite dimensional lattice to a completely disordered network. The ring-lattice does not show a small-world effect since the average shortest path grows faster than a logarithmic rate of increase with the number of vertices, but it has strong local clustering. When the edges are rewired, Watts and Strogatz noticed that replacing a few long-distance connections hugely reduced the network's average shortest path and, as a result, a small-world effect appears. When $p_{sw} = 1$, the network becomes completely disordered where local clustering is no longer present and the average shortest path is small. Watts and Strogatz showed, by numerical simulation, that there is a relatively large p_{sw} interval in between the two extremes, for which the model exhibits both low path lengths and clustering (Fig. 6.2).

Newman (2001, 2004) provides an excellent overview of the analysis on the topology of small-world network structures, highlighting key organizing principles that guide ties among the nodes in the network. According to Moody (2004), an archetypal small-world network will have many distinct clusters, connected to each other by a small number of ties. An analysis dealing with the dynamics of co-authorship publication networks in Slovenian sociology (Mali et al. 2010) showed that, to some extent, they conform to the small-world network structure: there are groups of sociologists that are very connected inside small groups but connected with others in non-systematic ways. Further results, obtained by using the blockmodeling approach, pointed to a publication strategy of those sociologists in Slovenia who are included in these small-world structures and are more oriented
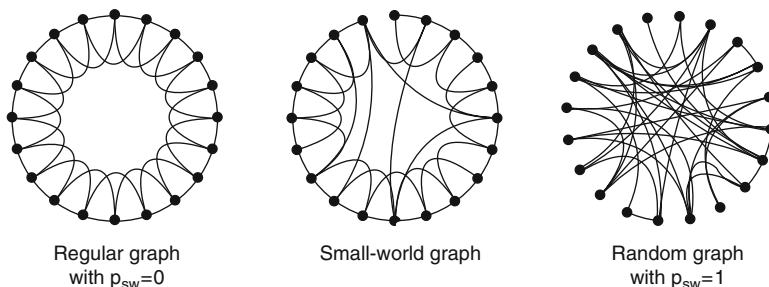

Regular graph
with $p_{sw}=0$ Small-world graph   Random graph
                                                               with $p_{sw}=1$

Fig. 6.2 Small-world structure simulation with different levels of randomness

to parochial scientific reports or publications in Slovene. Consistent with this, they publish less in the international peer-reviewed journals than the sociologists outside this small-world structure. The results of these empirical analyses of Slovenian sociologists suggest that the presence of a too 'closed' and dense co-authorship network in science can have negative effects on the international orientations of scientists in a small scientific community. This implies that, for scientific performance and scientific excellence, it is much more important to have 'open' networks that have many structural holes (gaps between actors that create opportunities for brokerage). This is especially important for linking micro-level interactions (cooperation inside internal scientific organizations) to macro-level patterns (cooperation in the international scientific community). Burt provided evidence suggesting that new ideas in society emerge from selection and synthesis processes that operate across structural holes between groups. Positive performance evaluations and good ideas are disproportionately in the hands of people whose networks span structural holes. The 'between-group brokers' are more likely to have ideas viewed as valuable (Burt 2004) within the community.

6.4.4.3   The Preferential Attachment Model

The scale-free network structure, in one version or another, corresponds fairly closely to the sociological model of cumulative advantage in science. The first systematic representation of this model was provided by Merton (1973). Following Merton, there was a research stream in the literature that invoked the idea of cumulative advantage as a central explanatory principle for the social stratification of science. Merton's studies were concerned with both organizational and functional aspects of science as an institution capable of self-regulation. This approach found its most significant (or at least most famous) expression in the description of the normative structure of science. Merton focused his attention on four institutional imperatives: universalism, communism, disinterestedness, and organized skepticism. Merton and other scholars working within institutional approaches (including Barber, Zuckerman, and Hagstrom) analyzed how norms regulate scientific activity. They studied the ways in which resources and rewards (including scientific prestige and opportunities to publish) are assigned and distributed within the scientific community (see, for example, Matthew 2005; Bucchi 2004).

The idea of cumulative advantages comes from the passage in Matthew's Gospel: "For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath." (Hence the term "the Matthew effect.") Translating the idea of cumulative advantage in science implies that those scientists who already occupy a position of excellence are rewarded far more than others in their field. Scientists who are rich in recognition find it easier to obtain additional recognition. In contrast, scientists who receive little recognition for their research efforts have reduced chances for future recognition.

Merton argued that cumulative advantage is a primary mechanism in modern science for the creation of scientific stars.[9]

A more quantitative and bibliometric basis for assessing the phenomenon of unequal distribution of publications (in connection with the unequal distribution of awards) in modern science has been provided also by Price (1976; 1963) in the form of his measure of scientific productivity. According to Price's law of scientific productivity, "...half of the scientific papers published in a given sector are signed by the square root of the total number of scientific authors in that field" (Price 1963, p. 67). This means that a relatively small number of highly productive researchers are responsible for most scientific publications. Price's law is founded on the same probabilistic basis as the earlier established Lotka Law,[10] the Bradford Law,[11] and Pareto and Zipf[12] distributions.

Both Price's law and the Matthew effect depict the scientific community as a structure characterized by marked inequality and a heavily pyramidal distribution of scientific rewards and publications. They are linked by the principle of preferential attachment which contains, for the case of scientific co-authorship networks, two generic aspects: (1) the continuous addition of new vertices into the network system and (2) preferential connectivity of new vertices. It means that a common feature of the models of scientific co-authorship networks, based on the rationale of preferential attachment, continuously expands by the addition of new vertices that are connected to the vertices already present in the networks. Additionally, in these models a new actor is, at best, most likely to be cast in a supporting role with more established and better-known actors. Further, no scientific field expands with

---

[9]Merton and his sociological followers (see Allison et al. 1982; Cole and Cole 1973) have analyzed several other similar mechanisms with regard to science networks, collaboration structures, and recognition in science:

1. The "halo effect" in science denotes the advantage of scientists in more favorable institutional locations.
2. The "Matilda effect" points to the discrimination against the participation of women in scientific activity.
3. The "gatekeeper" labels those scientists who can influence the distribution of resources such as research funds, teaching positions, or publishing opportunities because they occupy key decision-making positions within scientific institutions.
4. The idea of an "invisible college" was introduced on the basis of a seventeenth century expression denoting informal communities of researchers that cluster around specific projects or a research theme and that often turn out to be more influential in terms of knowledge production than formal communities (departments, research centers, scientific committees).

[10]Lotka's law states: The number of authors making n contributions is about $1/n^a$ of those making one contribution, where a is often about 2.

[11]Bradford's law states: Journals in a field can be divided into three parts: (1) a core of a few journals, (2) a second zone, with more journals, and (3) a third zone, with the bulk of journals. The number of journals in these three parts is $1 : n : n^2$.

[12]Zipf's law states: The probability of occurrence of words or other items starts high and tapers off. Thus, a few occur very often while many others occur rarely. The formal definition is: $P_n \sim 1/n^a$, where P_n is the frequency of occurrence of the nth ranked item and a is close to 1.

an endless growth of new vertices but is constrained by the operation of feedback effects.[13] It follows that there exist nodes, called "hubs" or "Angelpunkten oder Naben" (Berg 2005, p. 53), that acquire more links than another nodes. In such types of networks, preferential attachment and the system feedback dynamics play very important roles.

Crane (1972) provided an analysis of (global) scientific networks where informal members of scientific elites (in Moody's terminology, scientific stars) through whom the communication of scientific information both within scientific disciplines and across scientific disciplines is directed have the position of "hubs". Namely, they are central scientists in the network from where the information is transferred to all other scientists in the network. They also communicate intensively with each other. The idea of scientific networks with hubs can be used as a starting point to relate micro-level interactions (for example, in a local/national scientific community) to macro-level patterns (for example, the global scientific community). Through the informal groups of scientific elites, the small-scale interactions become translated into large-scale patterns. These large-scale patterns (international science) also have feedback effects on small groups (parochial/national science). The production and diffusion of the most creative and excellent scientific ideas in the world arise from the brokered networks (Granovetter 1973, p. 1360).

Albert and Barabási (2001) provide examples of many real-world networks whose degree distributions are far from a Poisson distribution. They showed that distributions can be approximated with a power-law function. They proposed a new evolving network model – PA or preferential attachment model (Barabási and Albert 1999). The model was presented as one that "shifts from modeling network topology to modeling the network assembly and evolution" (Albert and Barabási 2001). The idea behind the model was to capture the construction (development) of a network that could possibly explain the large number of observed power-law degree distributions in real networks. Before, there existed mostly network models with a fixed number of vertices among which links were added according to a particular procedure (process). Since real networks typically grow with the addition of new links and vertices that are not added randomly, Albert and Barabasi included the following ideas in their model.

The algorithmic statement of their model, given a set of vertices in a network, consists of the following two processes in a sequence of steps:

- At every time step, a new vertex v is added to the network.
- m_{ba} edges are created from the new vertex v to the vertices that are already in the network. These vertices are chosen with a probability proportional to their current

---

[13](Berg 2005, p. 54) points out that "the effect of the positive feed-backs, namely, the advantages of old nodes against new ones as well as the attractiveness of the already networked nodes for newly added ones are leading to the growth of networks based on the preferential attachment", ("...doch in einem bestimmten Bereich sind positiven Rueckkopplungen feststellbar. Beide Effekte zusammen, der Vorteil, den alte Knoten gegenueber neuen haben sowie die Attraktivitaet besonders vernetzter Knoten fuer neu hinzukommende, fuehren dazu, dass das Wachstum des Netzes einer bevorzugehenden Verbindungswahl folgt.")
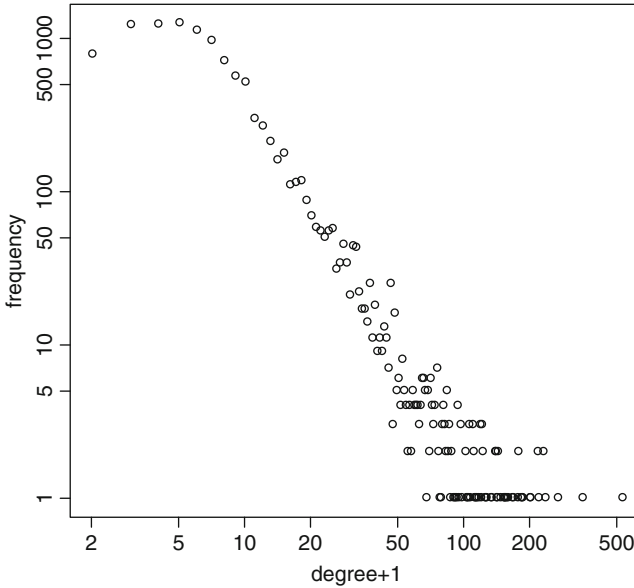
Fig. 6.3 Degree distribution in a co-authorship network of Slovenian researchers (Kronegger et al. 2011a) presented on a log-log scale. A large number of researchers with a small number of co-authors and a small number of researchers with a high number of co-authors indicates the existence of a preferential attachment mechanism in the process of network growth

degree. The probability of choosing vertex u can be written by $k_u / \sum_j k_j$ (where k_u represents the current degree of vertex u).

After t time steps, there are $t + m_0$ vertices in the network (where m_0 denotes the number of vertices at the beginning of the process) and $t m_{ba}$ edges. It was first shown with simulations that the degree distribution of the whole network resulting from the operation of this model follows a power-law distribution with an exponent $\gamma = 3$ (Fig. 6.3).

Such scale-free networks as these generated through the principle of preferential attachment, in addition to not having a Poisson distribution of links around nodes, also have the interesting property of being very resistant to random attack. Almost 80% of the links can be cut before a scale-free network is destroyed, while the corresponding percentage for an exponential network is less than 20%.

Many generalizations about preferential attachment models have been made (Albert and Barabási 2001; Newman 2003). Systematic divergence from the power-law distribution at small degrees can be seen in many real-world networks. Therefore, Pennock et al. (2002) proposed incorporating a mixture (weighted addition) of preferential attachment and random attachment in the model. A further refinement of this model, where a directed version of the model was taken into account, is implemented in Pajek (Batagelj and Mrvar 2010). There, at each step of the growth a new vertex is selected according to its weighted in-/outdegree and some uniform attachment.

Another generalization about both small-world and preferential attachment, developed for two-mode networks, comes from Latapy et al. (2008) who present a nice overview of method developments for two-mode networks. Opsahl (2010) provides another attempt to overcome the issues of higher clustering coefficients in projections of two-mode to one-mode networks by redefining both the global and local clustering coefficients so that they can be calculated directly for two-mode structures.

6.4.4.4 Applications Featuring Co-Authorship Networks

Newman (2001) showed that collaboration networks form small-worlds in which randomly chosen pairs of scientists are typically separated by only a short path of intermediate acquaintances. He further provided information on the distribution of the number of collaborators, demonstrated the presence of clustering in the networks, and highlighted the number of apparent differences in the patterns of collaboration between fields. Also, Newman (2004) used data from three bibliographic databases for biology, physics, and mathematics to construct networks in which the nodes were scientists. He used these networks to answer a broad variety of questions about collaboration patterns, how many papers did authors write and with how many people, what is the typical distance between scientists through the network, and how do patterns of collaboration vary between subjects and over time.

Barabási et al. (2002) analyzed co-authorship data from electronic databases containing all relevant journals in mathematics and neuroscience for the period between 1991 and 1998. They found that network evolution is governed by preferential attachment. However, contrary to their predictions, the average degree in the networks they analyzed increased, and the node separation decreased in time. They also proposed a model that captured the network's time evolution.

Moody (2004) made an important contribution by identifying several types of individual scientific collaboration behavior that leads to the development of co-authorship networks that resemble networks generated according to the principles of small-world and preferential attachment. Recently, several articles that test the principles of small-world and preferential attachment have been published. Some are based on local databases like the Slovenian *COBISS* (Mali et al. 2010; Perc 2010), while others use general databases like *Web of Science* (Wagner and Leydesdorff 2005; Tomassini and Luthi 2007).

6.4.4.5 Developments of Models for Longitudinal Network Data

After the pioneering work of Erdös and Rényi on random graphs, and after the first applications of graph theory appeared in the sociological community (de Sola Pool and Kochen 1978), one group within the scientific community moved *away* from the idea of merely reproducing some global properties of "real-world" network properties. Instead, they focused on an approach designed to include micro-mechanisms

that generate local changes in networks that also help account for the macro-structure of networks. Moreover, these efforts were designed to treat the micro-mechanisms as hypotheses that could be evaluated through statistical inference. The basics for these models of network change are (as already mentioned in the previous section) random graphs and random graph processes which incorporate the probabilistic uncertainty into the models. Uncertainty is present because there are many potential generators for observed graph structures, including co-authorship. From a methodological perspective, modeling the dynamics in social networks led to several obstacles. Probably the most persistent one was the interdependencies of the units comprising the networks. This problem remained untouched for almost 20 years. Indepth overview of approaches and methods to modeling network changes in time can be found in Frank (1991), Snijders (1996), and Snijders et al. (2010).

There are two distinct approaches to modeling network changes in time: models that implement change in discrete time steps, and more advanced models where time is modeled by continuous flows. Success in modeling change in social networks began in 1959 when Katz and Proctor showed that change in preferences for making ties in the network could be represented by a stationary, discrete-time Markov model. Of course, they assumed the independence of dyads within which all the modeling took place. In 1981, Holland and Leinhardt published a very influential article on log-linear models of network change which initiated a vigorous research stream devoted to the development of a broad class of models. One basic model, called p_1, was developed by Fienberg and Wasserman (1981) and Wasserman and Weaver (1985). Authors also provided efficient algorithms to find the maximum-likelihood estimators of parameters defining appropriate probability functions. Fienberg et al. (1985) showed how to handle social network data with the Holland-Leinhardt model and its extensions in contingency tables by using basic log-linear models. The longitudinal dimension to the log-linear approach was added by Wasserman (1987) and Wasserman and Iacobucci (1988).

Conditionally uniform models (Holland and Leinhardt 1975) are often used for modeling directed graphs where the probability distribution for forming new ties is uniform, conditional upon a certain set of attributes. In these models, the conditional statistics are defined by attribute variables and contain the most relevant effects of the studied phenomena, while the rest is explained by random factors. Conditionally uniform models become very complicated when more informative conditioning on attribute variables is included into the model. Such models for longitudinal binary network data at 2 time points – conditional upon the entire network at the first time point, and upon the numbers of newly formed and dissolved ties for each actor – were developed by Snijders (1990). The idea of conditioning the changes in the network on the first measured network resolves most of the unexplained factors that determined the development of network before its first measurement.

Modeling changes in continuous time with Markov chains was adapted by Coleman (1964) to tackle some classical sociological problems. Holland and Leinhardt (1977) extended this idea to model networks of interpersonal affect between actors. They developed a valued Markov chain approach to model the process by which social structure based on affect influenced individual behavior.

The basic assumptions underlying the use of the continuous time Markov chain model are:

1. Between the observation moments, time runs continuously. Changes can be made (but are likely to be unobserved) at any moment, t.
2. The network $X(t)$ is the outcome of a Markov process.
3. At each single moment, only one relational tie or variable attribute may change.

Wasserman (1978, 1980a,b) continued this approach and provided estimators for parameters of various models. He started with a simple model of reciprocity in directed graphs, but without complicated dependencies between ties such as those generated by transitive closure.

The breakthrough in modeling the dynamics in social networks was the relaxation of the assumption of conditional independence between dyads (Mayer 1984). This was an important step since most sociological theories assume at least some kind of dependence structure between dyads. Another important step came in the form of dropping the stationarity assumption (Leenders 1995). Leenders also developed a mechanism to allow changing rates for all dyads to be dependent on arbitrary covariates, with the assumption that these remain constant between the observations.

In recent years, these models became known as stochastic actor-oriented models which have been developed to consider a variety of micro-mechanisms for generating network structure. These models are based on an assumption that each actor has his/her own goals which he/she tries to advance in accordance with his/her constraints and possibilities. Snijders (1995) referred to this approach as 'methodological individualism' where the driving force behind the network dynamics comes in the form of actions by actors.

Each attempt to model specific sociological problems or theories produced a new mathematical model that filled the gaps along the way to obtaining a better representation of reality. Yet an important feature still had to be addressed because most of these models lacked an explicit estimation theory.

The first models addressed some basic questions. A baseline of development can be followed through the work of several authors. Jackson and Wolinsky (1996) presented a model where the benefits and costs of ties affected the evolutionary trajectories of networks and the form of equilibrium structures. Hummon (2000) constructed actor-oriented simulation models of 'Jackson and Wolinsky actors' to study temporal network dynamics. He specified choices under four combinations of tie formation and deletion rules: unilateral and mutual tie formation, and unilateral and mutual tie deletion. This process generated eight types of networks: Null, near-Null, Star, near-Star, Shared, near-Shared, Complete and near-Complete as equilibrium structures. Doreian (2006) provided a formal proof via exhaustive examinations of the structures identified by Hummon (but only for tiny networks), and this line of work was extended by Xie and Cui (2008a,b). In another line of development, Marsili et al. (2004) presented a simple model using the creation of links to friends of friends, a mechanism that was introduced by Vázquez (2003) in the context of growing networks. This model is similar to the one proposed by

Davidsen et al. (2002) which explained the emergence of the small-world property in some social networks.

In the model of Skyrms and Pemantle (2000), individual agents begin to interact at random, with the interactions modeled as games. The game payoffs determine which interactions are reinforced, and network structures emerge as a consequence of the dynamics of the agents' learning behavior.

More complex network dynamic models with larger but still quite restricted numbers of tendencies were presented by Jin et al. (2001). They propose some simple models for the growth of social networks based on three general principles: (i) meetings take place between pairs of individuals at a rate that is high if a pair has one or more mutual friends and low otherwise; (ii) acquaintances between pairs of individuals who rarely meet decay over time; (iii) there is an upper limit on the number of friendships an individual can maintain. Their models incorporate all of these principles and reproduce many of the features of real social networks, including high levels of clustering or network transitivity and strong community structure in which individuals have more links to others within their community than they have to individuals from other communities. The important feature of their models is the inclusion of a time scale on which people make and break social connections.

6.4.4.6 Simulation Investigation for Empirical Network Analysis – Siena

The problem of inference in modeling dynamics of social networks on the basis of the observed longitudinal data was addressed by Snijders (1996) and extended further by Snijders et al. (2010). These models are based on longitudinal data and include representations of network dynamics as being driven by many different tendencies. These include micro-mechanisms, which have been theoretically derived and/or empirically established in earlier research, and which may well operate simultaneously. One of the most important characteristics of these models is the evaluation of their results within the paradigm of statistical inference, which makes them suitable for testing hypotheses and estimating tendencies that drive tie formation and dissolution at the level of individual units using reciprocity, transitivity, homophily, etc.

The model assumptions are:

- The model is basically defined for directed relations. In the case of undirected networks (e.g., co-authorship networks) the tie formation is additionally modeled using different mechanisms (e.g., a unilateral forcing model, unilateral initiative, and reciprocal confirmation, etc.)
- The network is observed in 2 or more discrete timepoints. But the underlying time parameter in the model is continuous.
- Changes in the network are outcomes of a Markov process, which means that the change in the network from one state in time point t_i to new state in time point

t_{i+1} is conditioned only to the state of the network in time point t_i. The process does not take into account any other historical events.

- The actors control their ties, which means that changes in ties are made by actors who send the tie on the basis of their and others' attributes, their position in the network, and their perceptions about the rest of the network. Regarding the last, it is assumed that actors have full information about the network and the other actors.
- At any given moment, only one probabilistically selected actor may get the opportunity to change only one tie.

The actor-based process is decomposed into two stochastic sub-processes:

1. The change-opportunity process models the frequency of the tie changes by actors. The opportunity to change the tie depends on the network locations of the actor (e.g., his or her centrality) and on actor covariates (e.g., gender or age).
2. The change-determination process models the change of the tie when an actor gets an opportunity to make a change. The change of the tie can be made with equal probabilities or with probabilities depending on attributes or network positions. Perceived attributes and position (the environment) of the actor is included into the actor's objective function, which expresses how likely it is for the actor to change his or her network environment in specific way (i.e., initiate, withdraw tie, or keep the present situation).

To use this model with observed data means that parameters have to be estimated by some statistical procedure. Since the model is too complicated for classical estimation methods such as maximum likelihood, Snijders (1996, 2001) proposed a procedure using the method of moments implemented by a computer simulation of the network change process. The procedure he proposed uses the first observation of the network as the (unmodeled) starting point of the simulations. This implies the estimation procedure is conditioned on the first observed network of a series of observations of that network.

The limitation of such models is that they are limited to a predetermined and rather small number of actors (between 100 and 200 actors) and do not directly consider the mechanisms of network growth. The methods and algorithms developed by Snijders et al. (2008) are implemented in the computer package SIENA.

Stochastic actor-based modeling of network dynamics was initially developed for modeling the change in directed networks. The undirected networks such as co-authorship networks are a special case where reciprocity cannot be used as a mechanism of network change. Although several articles have been published using SIENA models, to our knowledge, only Kronegger et al. (2011b) dealt with undirected networks to study the dynamics of co-authorship networks of Slovenian researchers working in physics, mathematics, biotechnology, and sociology in the time period from 1991 to 2005. In their study, they operationalized the modeling of global network parameters used in the preferential attachment and the small-world models with stochastic actor-oriented modeling.

6.5   Summary

Access to bibliographic databases and the availability of powerful quantitative social network approaches increased the number of studies of co-authorship networks in different scientific fields. There are several classification schemes for analytical approaches to analyzing the dynamics of co-authorship networks. We decided to classify them according to the types of models. The first type of model provides the basic analysis of whole co-authorship network properties. Such network characteristics are degrees, clustering coefficients, and density. The usual statistical approach used in these models is time-series analysis of listed properties.

Deterministic models (the second type) and stochastic models (the third type) are usually used to analyze actor-based co-authorship networks and attribute characteristics. To study the structure within the co-authorship networks, blockmodeling approaches are recommended. To model dynamic co-authorship networks, several approaches can be used according to the chosen level of analysis. Models on the macro level (whole network level) were mostly developed by mathematicians and physicists. These are models of "real-world" networks, small-world models, and preferential attachment models. The alternative stochastic actor-based model (implemented in SIENA) was developed by social scientists and statisticians. This model focuses on single units and on dyads. This powerful model studies network change in time as the result of micro-mechanisms for generating the network structure.

There are several indicators that show a huge development of analytical approaches to studying social networks through time. The powerful stochastic actor-based networking model has one disadvantage in that it can only be used to analyze a few hundred units in the network. Therefore, there is a need for similar models to analyze large networks.

Key points

Modeling of co-authorship networks can be approached in terms of the different perspectives and goals that have been outlined in this chapter. As a partial summary, the following items are important:

1. Level of the analysis: the macro level (whole network) or the micro level (unit). Which one is used depends on the goal(s) of the study. There are the following three variants:

   a. Describing the topology of the macro structure
   b. Understanding the micro-level changes at the actor level
   c. Coupling the micro-level processes to the generation of the network's macro structure.

2. Size of the network: some models can process only a limited number of units (e.g. stochastic actor-based modeling and direct blockmodeling),

while others can handle large networks (e.g., preferential attachment, the small-world model, and indirect blockmodeling).

3. Discrete-time models (e.g., blockmodeling) or continuous-time models (e.g., stochastic actor based modeling).

4. The analysis of the evolution of co-authorship networks only (e.g., small-world model, preferential attachment, blockmodel) or including external characteristic of network (e.g., scientific field) and/or actor attributes (e.g., age or gender of researcher) using modeling approaches (e.g., stochastic actor based modeling).

5. Needs of graphical representation of co-authorship network evolution (e.g., preferential attachment, blockmodeling, multidimensional scaling).

References

Albert R, Barabási AL (2001) Statistical mechanics of complex networks. Rev Mod Phys 74(1):47–97 (DOI: 10.1103/RevModPhys.74.47), also available as arXiv preprint arXiv:cond-mat/0106096

Allison PD, Long JS, Krauze TK (1982) Cumulative advantage and inequality in science. Am Sociol Rev 47(5):615–625 (DOI: 10.1146/annurev.soc.32.061604.123127, stable JSTOR URL: http://www.jstor.org/stable/209516)

Andrade HB, de los Reyes Lopez E, Martin TB (2009) Dimensions of scientific collaboration and its contribution to the academic research groups' scientific quality. Res Eval 18(4):301–311 (DOI: 10.3152/095820209X451041)

Babchuk N, Keith B, Peters G (1999) Collaboration in sociology and other scientific disciplines: A comparative trend analysis of scholarship in the social, physical, and mathematical sciences. Am Sociol 30(3):5–21 (DOI: 10.1007/s12108-999-1007-5, stable JSTOR URL: http://www.jstor.org/stable/27698918)

Barabási AL (2002) Linked: The new science of networks. Perseus, Cambridge, MA

Barabási AL, Albert R (1999) Emergence of scaling in random networks. Science 286(5439):509–512 (DOI: 10.1126/science.286.5439.509), also available as arXiv preprint arXiv:cond-mat/9910332

Barabási AL, Jeong H, Néda Z, Ravasz E, Schubert A, Vicsek T (2002) Evolution of the social network of scientific collaborations. Physica A 311(3–4):590–614 (DOI:10.1016/S0378-4371(02)00736-7), also available as arXiv preprint arXiv:cond–mat/0104162

Batagelj V, Mrvar A (2010) Pajek – analysis and visualization of large networks: Reference manual: List of commands with short explanations: version 2.00, 1st edn. University of Ljubljana, Ljubljana, available online at the URL: http://vlado.fmf.uni-lj.si/pub/networks/pajek/doc/pajekman.pdf

Batagelj V, Doreian P, Ferligoj A (1992a) An optimizational approach to regular equivalence. Soc Networks 14(1–2):121–135 (DOI: 10.1016/0378-8733(92)90016-Z)

Batagelj V, Ferligoj A, Doreian P (1992b) Direct and indirect methods for structural equivalence. Soc Networks 14(1–2):63–90 (DOI: 10.1016/0378-8733(92)90014-X)

Batagelj V, Ferligoj A, Doreian P (1998) Fitting pre-specified blockmodels. In: Hayashi N, Jajima K, Bock HH, Ohsumi N, Tanaka Y, Baba Y (eds) Data science, classification, and related methods: Proceedings of the fifth conference of the International Federation of Classification

Societies (IFCS-96), Kobe, Japan, March 27–30, 1996. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Tokyo, pp 199–206

Bayer AE, Smart JC (1991) Career publication patterns and collaborative "styles" in american academic science. J High Educ 62(6):613–636 (DOI: http://dx.doi.org/10.2307/1982193, stable JSTOR URL: http://www.jstor.org/stable/1982193)

Berg C (2005) Vernetzung als Syndrom. Risiken und Chancen von Vernetzungsprozessen für eine nachhaltige Entwicklung. Campus: Forschung, vol 883. Campus Verlag, Frankfurt am Main

Börner K, Boyack KW, Milojević S, Morris S (2012) Basic model types, key definitions, and a general framework for the comparison of process models. In: Scharnhorst A, Börner K, van den Besselaar P (eds) Models of science dynamics. Springer, Berlin, Ch 1, pp 3–22

Bonacich P (2004) The invasion of the physicists. Soc Networks 26(3):285–288 (DOI:10.1016/j.socnet.2004.06.002)

Bordons M, Morillo F, Gómez I (2004) Analysis of cross-disciplinary research through bibliometric tools. In: Moed HF, Glänzel W, Schmoch U (eds) Handbook of quantitative science and technology research: The use of patent and publication statistics in studies on S&T systems. Kluwer, Dordrecht, pp 437–456 (DOI: 10.1007/1-4020-2755-9_20)

Braun T, Schubert A (2003) A quantitative view on the coming of age of interdisciplinarity in the sciences 1980–1999. Scientometrics 58(1):183–189 (DOI: 10.1023/A:1025439910278)

Braun-Munzinger P (2009) Teilchenphysik und Teilgebiete der Kernphysik: Das Publikationsverhalten in Großkollaborationen. In: Publikationsverhalten in unterschiedlichen wissenschaftlichen Disziplinen: Beiträge zur Beurteilung von Forschungsleistungen, 2nd ext. edn. Diskussionspapiere der Alexander von Humboldt-Stiftung, No 12. Alexander von Humboldt Foundation/Stiftung, Bonn, pp 92–94, available online at the URL: http://www.humboldt-foundation.de/pls/web/docs/F13905/12_disk_papier_publikationsverhalten2_kompr.pdf

Bucchi M (2004) Science in society: An introduction to social studies of science. Routledge, London. Revised and expanded ed. of: Bucchi M (2002) Scienza e Società: Introduzione alla sociologia della scienza. Il Mulino, Bologna

Burt RS (2004) Structural holes and good ideas. Am J Sociol 110(2):349–399 (DOI:10.1086/421787, stable JSTOR URL: http://www.jstor.org/stable/10.1086/421787)

Buter RK, Noyons, van Raan AFJ (2010) Identification of converging research areas using publication and citation data. Res Eval 19(1):19–27 (DOI: 10.3152/095820210X492503)

Cole S, Cole JR (1973) Social stratification in science. University of Chicago Press, Chicago, IL

Coleman JS (1964) An introduction to mathematical sociology. Free Press, New York, NY; Collier-Macmillan, London

Crane D (1972) Invisible colleges: Diffusion of knowledge in scientific communities. University of Chicago Press, Chicago, IL

Davidsen J, Ebel H, Bornholdt S (2002) Emergence of a small world from local interactions: Modeling acquaintance networks. Phys Rev Letters 88(12):128701, 4 pp (DOI: 10.1103/PhysRevLett.88.128701), also available as arXiv preprint arXiv:cond-mat/0108302

de Haan J (1997) Authorship patterns in Dutch sociology. Scientometrics 39(2):197–208 (DOI:10.1007/BF02457448)

de Sola Pool I, Kochen M (1978) Contacts and influence. Soc Networks 1(1):5–51 (DOI:10.1016/0378-8733(78)90011-4)

Doreian P (2006) Actor network utilities and network evolution. Soc Networks 28(2):137–164 (DOI: 10.1016/j.socnet.2005.05.002)

Doreian P, Stokman FN (eds) (1997) Evolution of social networks. Gordon and Breach, Amsterdam; Routledge, London

Doreian P, Batagelj V, Ferligoj A (2005) Generalized blockmodeling. Structural Analysis in the Social Sciences, vol 25. Cambridge University Press, Cambridge (DOI:10.1017/CBO9780511584176)

Erdös P, Rényi A (1959) On random graphs I. Publicationes Mathematicae (Debrecen) 6:290–297, available online at the site: http://www.renyi.hu/~p_erdos/Erdos.html at the URL: http://www.renyi.hu/~p_erdos/1959-11.pdf. Reprinted in: Turán P (ed) (1976) Rényi Alfréd Válogatott Munkái/Selected Papers of Alfréd Rényi, (volume 2, 1956–1961. Akadémiai Kiadó, Budapest, pp 308–315

Erten C, Harding PJ, Kobourov SG, Wampler K, Yee G (2004) GraphAEL: Graph animations with evolving layouts. In: Liotta G (ed) Graph drawing: 11th international symposium, GD 2003, Perugia, Italy, September 21–24, 2003: Revised papers. Lecture Notes in Computer Science, vol 2912. Springer, Berlin, pp 98–110 (DOI: 10.1007/978-3-540-24595-7_9)

Etzkowitz H, Leydesdorff L (2001) Universities in the global knowledge economy. In: Etzkowitz H, Leydesdorff L (eds) Universities and the global knowledge economy: A triple helix of university-industry-government relations. Science, Technology and the International Political Economy. Pinter, London, pp 1–11

Fienberg SE, Wasserman SS (1981) Categorical data analysis of single sociometric relations. In: Leinhardt S (ed) Sociological methodology 1981. Jossey-Brass, San Francisco, CA, pp 156–192 (DOI: 10.2307/270741 stable JSTOR URL: http://www.jstor.org/stable/270741).

Fienberg SE, Meyer MM, Wasserman S (1985) Statistical analysis of multiple sociometric relations. J Am Stat Assoc 80(389):51–67 (DOI: 10.2307/2288040, stable JSTOR URL: http://www.jstor.org/stable/2288040)

Frank O (1991) Statistical analysis of change in networks. Statistica Neederlandica 45(3):283–293 (DOI: 10.1111/j.1467-9574.1991.tb01310.x)

Frank O, Strauss D (1986) Markov graphs. J Am Stat Assoc 81(395):832–842 (DOI:10.2307/2289017, stable JSTOR URL: http://www.jstor.org/stable/2289017)

Freeman LC (2004) The development of social network analysis: A study in the sociology of science. EP Empirical Press, Vancouver, BC; BookSurge, North Charleston, SC

Fuchs S (1992) The professional quest for truth: A social theory of science and knowledge. SUNY Series in Science, Technology, and Society. State University of New York Press, Albany, NY

Gansner ER, Koren Y, North SC (2005) Graph drawing by stress majorization. In: Pach J (ed) Graph drawing: 12th international symposium, GD 2004, New York, NY, USA, September 29–October 2, 2004: Revised selected papers. Lecture Notes in Computer Science, vol 3383. Springer, Berlin, pp 239–250 (DOI: 10.1007/978-3-540-31843-9_25)

Gibbons M, Limoges C, Nowotny H, Schwartzman S, Scott P, Trow M (1994) The new production of knowledge: The dynamics of science and research in contemporary societies. Sage Publications, London

Glänzel W, Schubert A (2004) Analysing scientific networks through co-authorship. In: Moed HF, Glänzel W, Schmoch U (eds) Handbook of quantitative science and technology research: The use of patent and publication statistics in studies on S&T systems. Kluwer, Dordrecht, pp 257–276 (DOI: 10.1007/1-4020-2755-9_20)

Glänzel W, Schubert A (2005) Domesticity and internationality in co-authorship, references and citations. Scientometrics 65(3):323–342 (DOI: 10.1007/s11192-005-0277-0)

Glänzel W, Schubert A, Czerwon HJ (1999) A bibliometric analysis of international scientific cooperation of the European Union (1985–1995). Scientometrics 45(2):185–202 (DOI: 10.1007/BF02458432)

Gossart C, Ozman M (2009) Co-authorship networks in social sciences: The case of Turkey. Scientometrics 78(2):323–345 (DOI: 10.1007/s11192-007-1963-x)

Granovetter MS (1973) The strength of weak ties. Am J Sociol 78(6):1360–1380 (DOI:http://dx.doi.org/10.2307/202051, stable JSTOR URL: http://www.jstor.org/stable/2776392)

Handcock MS, Hunter DR, Butts CT, Goodreau SM, Morris M (2003) Statnet: Software tools for the statistical modeling of network data, version 2.0. University of Washington, Center for Studies in Demography and Ecology, Seattle, WA, available online at the URL: http://statnetproject.org

Handcock MS, Raftery AE, Tantrum JM (2007) Model-based clustering for social networks. J R Stat Soc Ser A Stat Soc 170(2):301–354 (DOI: 10.1111/j.1467-985X.2007.00471.x), also available as University of Washington, Center for Statistics and the Social Sciences Working Paper no. 46 at the URL: http://www.csss.washington.edu/Papers/wp46.pdf

Havemann F, Heinz M, Kretschmer H (2006) Collaboration and distances between German immunological institutes – a trend analysis. J Biomed Discov Collab 1(6), 7 pp (DOI:10.1186/1747-5333-1-6)

Hoekman J, Koen F, Tijssen R (2010) Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. Res Policy 39(5):662–673 (DOI:10.1016/j.respol.2010.01.012)

Holland PW, Leinhardt S (1975) The statistical analysis of local structure in social networks. In: Heise DR (ed) Sociological methodology 1976. Jossey-Brass, San Francisco, CA, pp 1–45 (DOI: 10.2307/270703, stable JSTOR URL: http://www.jstor.org/stable/270703).

Holland PW, Leinhardt S (1977) A dynamic model for social networks. J Math Sociol 5(1):5–20 (DOI: 10.1080/0022250X.1977.9989862)

Holland PW, Leinhardt S (1981) An exponential family of probability distributions for directed graphs. J Am Stat Assoc 76(373):33–50 (DOI: 10.2307/2287037), stable JSTOR URL: http://www.jstor.org/stable/2287037)

Hornbostel S (1997) Bewertungen in der Wissenschaft. Wissenschaftsindikatoren. Westdeutscher Verlag, Opladen

Hummon NP (2000) Utility and dynamic social networks. Soc Networks 22(3):221–249 (DOI: 10.1016/S0378-8733(00)00024-1)

Hummon NP, Carley K (1993) Social networks as normal science. Soc Networks 15(1):71–106 (DOI: 10.1016/0378-8733(93)90022-D)

Hummon NP, Doreian P (1989) Connectivity in a citation network: The development of DNA theory. Soc Networks 11(1):39–63 (DOI: 10.1016/0378-8733(89)90017-8)

Hummon NP, Doreian P (1990) Computational methods for social network analysis. Soc Networks 12(4):273–288 (DOI: 0378-8733(90)90011-W)

Jackson MO, Wolinsky A (1996) A strategic model of social and economic networks. J Econ Theory 71(1):44–74 (DOI: 10.1006/jeth.1996.0108)

Jin EM, Girvan M, Newman MEJ (2001) Structure of growing social networks. Phys Rev E 64(4):046132, 8 pp (DOI: 10.1103/PhysRevE.64.046132)

Katz JS (1994) Geographical proximity and scientific collaboration. Scientometrics 31(1):31–43 (DOI: 10.1007/BF02018100)

Katz JS, Martin BR (1997) What is research collaboration? Res Policy 26(1):1–18 (DOI:10.1016/S0048-7333(96)00917-1)

Katz L, Proctor C (1959) The concept of configuration of interpersonal relations in a group as a time-dependent stochastic process. Psychometrika 24(4):317–327 (DOI:10.1007/BF02289814)

Kejžar N (2007) Evolution of networks. PhD thesis. University of Ljubljana, Faculty of Mathematics and Physics, Ljubljana

Kronegger L, Ferligoj A, Doreian P (2011a) On the dynamics of national scientific systems. Quality and Quantity 45(5):989–1015 (DOI: 10.1007/s11135-011-9484-3)

Kronegger L, Mali F, Ferligoj A, Doreian P (2011b) Collaboration structures in Slovenian scientific communities. Scientometrics (online first) (DOI: 10.1007/s11192-011-0493-8)

Kuhn TS (1996) The structure of scientific revolutions, 3rd edn. University of Chicago Press, Chicago, IL

Laband DN, Tollison RD (2000) Intellectual collaboration. J Polit Econ 108(3):632–662 (DOI: 10.1086/262132, stable JSTOR URL: http://www.jstor.org/stable/10.1086/262132)

Lambiotte R, Panzarasa P (2009) Communities, knowledge creation, and information diffusion. J Informet 3(3):180–190 (DOI: 10.1016/j.joi.2009.03.007), also available as arXiv preprint arXiv:0909.2481

Latapy M, Magnien C, Vecchio ND (2008) Basic notions for the analysis of large two-mode networks. Soc Networks 30(1):31–48 (DOI: 10.1016/j.socnet.2007.04.006)

Laudel G (2002) What do we measure by co-authorship? Res Eval 11(1):3–15 (DOI:10.3152/147154402781776961)

Leenders RTAJ (1995) Models for network dynamics: A Markovian framework. J Math Sociol 20(1):1–21 (DOI: 10.1080/0022250X.1995.9990149)

Leydesdorff L (2010) Eugene Garfield and algorithmic historiography: Co-words, co-authors, and journal names. Ann Libr Inf Stud 57(3):248–260, available online at the URL: http://nopr.niscair.res.in/handle/123456789/10246, also available as arXiv preprint arXiv:1005.5444. The

article is part of: Arunachalam S (ed) Special Issue: A tribute to Eugene Garfield, Information Scientist Extraordinaire, on his 85th birthday. Ann Libr Inf Stud 57(3):173–326

Leydesdorff L, Schank T (2008) Dynamic animations of journal maps: Indicators of structural changes and interdisciplinary developments. J Am Soc Inf Sci Technol 59(11):1532–2890 (DOI: 10.1002/asi.20891), also available as arXiv preprint arXiv:0911.1437

Lorrain F, White HC (1971) Structural equivalence of individuals in social networks. J Math Sociol 1(1):49–80 (DOI: 10.1080/0022250X.1971.9989788)

Lowrie A, McKnight PJ (2004) Academic research networks: A key to enhancing scholarly standing. Eur Manag J 22(4):345–360 (DOI: 10.1016/j.emj.2004.06.011)

Mali F, Kronegger L, Ferligoj A (2010) Co-authorship trends and collaboration patterns in the Slovenian sociological community. Corvinus J Sociol Soc Policy 1(2):29–50, available online at the URL: http://cjssp.uni-corvinus.hu/index.php/cjssp/article/view/24

Marsili M, Vega-Redondo F, Slanina F (2004) The rise and fall of a networked society: A formal model. Proc Natl Acad Sci USA 101(6):1439–1442 (DOI: 10.1073/pnas.0305684101, stable JSTOR URL: http://www.jstor.org/stable/3371207)

Matthew D (2005) Science in society. Palgrave Macmillan, Basingstoke

Mayer T (1984) Parties and networks: Stochastic methods for relationship networks. J Math Sociol 10(1):51–103 (DOI: 10.1080/0022250X.1984.9989958)

Melin G, Persson O (1996) Studying research collaboration using co-authorship. Scientometrics 36(3):363–377 (DOI: 10.1007/BF02129600)

Merton RK (1973) The sociology of science: Theoretical and empirical investigations. University of Chicago Press, Chicago, IL

Molloy M, Reed B (1995) A critical point for random graphs with a given degree sequence. Random Structures Algorithms 6(2–3):161–179 (DOI: 10.1002/rsa.3240060204)

Moody J (2004) The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. Am Sociol Rev 69(2):213–238 (DOI: 10.1177/000312240406900204, stable JSTOR URL: http://www.jstor.org/stable/3593085)

Mulkay M (1975) Three models of scientific development. Sociol Rev 23(3):509–526

Munshi UM, Pant P (2004) Regional, national, international coauthorship and subauthorship collaboration and author productivity as manifested in the scholarly journal literature of nuclear science. In: Kretschmer H, Singh Y, Kundra R (eds) International Workshop on Webometrics, Informetrics and Scientometrics (WIS 2004), Roorkee, March 2–5, 2004. Indian Institute of Technology, Roorkee, pp 273–290, available online at the URL: http://hdl.handle.net/123456789/461

Newman MEJ (2000) Models of the small world. J Stat Phys 101(3–4):819–841 (DOI:10.1023/A:1026485807148), also available as arXiv preprint arXiv:cond-mat/0001118

Newman MEJ (2001) The structure of scientific collaboration networks. Proc Natl Acad Sci USA 98(2):404–409 (DOI: 10.1073/pnas.021544898, stable JSTOR URL: http://www.jstor.org/stable/3054694), also available as arXiv preprint arXiv:cond-mat/0007214

Newman MEJ (2003) The structure and function of complex networks. SIAM Rev 45(2):167–256 (DOI: 10.1137/S003614450342480), also available as arXiv preprint arXiv:cond-mat/0303516

Newman MEJ (2004) Coauthorship networks and patterns of scientific collaboration. Proc Natl Acad Sci USA 101(Suppl 1):5200–5205 (DOI: 10.1073/pnas.0307545100). The article is part of the special issue: (2004) Mapping knowledge domains: Arthur M. Sackler Colloquium of the National Academy of Sciences. Proc Natl Acad Sci USA 101(Suppl 1):5183–5310

Nooraie RY, Kamrani MA, Hanneman R, Etemadi A (2008) Association between co-authorship network and scientific productivity and impact indicators in academic medical research centers: A case study in Iran. Health Res Policy Syst 6(9), 8 pp (DOI: 10.1186/1478-4505-6-9)

Opsahl T (2010) Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. Arxiv preprint arXiv:1006.0887

Pattison P, Wasserman S (1999) Logit models and logistic regressions for social networks: II. Multivariate relations. Br J Math Stat Psychol 52(2):169–193 (DOI: 10.1348/000711099159053)

Pennock DM, Flake GW, Lawrence S, Glover EJ, Giles CL (2002) Winners don't take all: Characterizing the competition for links on the web. Proc Natl Acad Sci USA 99(8):5207–5211 (DOI: 10.1073/pnas.032085699, stable JSTOR URL: http://www.jstor.org/stable/3058470)

Perc M (2010) Growth and structure of Slovenia's scientific collaboration network. J Informetr 4(4):475–482 (DOI: 10.1016/j.joi.2010.04.003), also available as arXiv preprint arXiv:1004.4824

Powell WW, White DR, Koput KW, Owen-Smith J (2005) Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences. Am J Sociol 110(4):1132–1205 (DOI: 10.1086/421508, stable JSTOR URL: http://www.jstor.org/stable/10.1086/421508)

Price DJ de Solla (1963) Little science, big science. Columbia University Press, New York, NY

Price DJ de Solla (1976) A general theory of bibliometric and other cumulative advantage processes. J Am Soc Inf Sci 27(5):292–306 (DOI: 10.1002/asi.4630270505), also available online at the URL: http://www.asis.org/Publications/JASIS/Best_Jasist/1976pricejasistarticle.pdf

Price DJ de Solla, Beaver D (1966) Collaboration in an invisible college. Am Psychologist 21(11):1011–1018 (DOI: 10.1037/h0024051), also available online at the URL: http://www.sciencedirect.com/science/article/pii/S0003066X07631281

Qin J, Lancaster FW, Allen B (1997) Types and levels of collaboration in interdisciplinary research in the sciences. J Am Soc Inf Sci 48(10):893–916 (DOI: 10.1002/(SICI)1097-4571(199710)48:10¡893::AID-ASI5¿3.0.CO;2-X)

Richardson MW (1938) Multidimensional psychophysics. Psychol Bull 35:659–660

Rigby J (2005) Handcrafted by 16 men: The impact of single and multiple authorship in collaborative research networks. Res Eval 14(3):199–206 (DOI: 10.3152/147154405781776148)

Rogers J, Bozeman B, Chompalov I (2001) Obstacles and opportunities in the application of network analysis to the evaluation on R&D. Res Eval 10(3):161–172 (DOI:10.3152/147154401781777033)

Said YH, Wegman EJ, Sharabati WK, Rigsby JT (2008) Social networks of author-coauthor relationships. Comput Stat Data Anal 52(4):2177–2184 (DOI: 10.1016/j.csda.2007.07.021)

Sailer L (1978) Structural equivalence: Meaning and definition, computation and application. Soc Networks 1(1):73–90 (DOI: 10.1016/0378-8733(78)90014-X)

Schott T (1993) World science: Globalization of institutions and participation. Sci Technol Human Values 18(2):196–208 (DOI: 10.1177/016224399301800205, stable JSTOR URL: http://www.jstor.org/stable/690067)

Sigogneau A, Malagutti O, Crance M, Bauin S (2005) Cross-disciplinary research: Co-evaluation and co-publication practices of the CNRS laboratories. Res Eval 14(2):165–176 (DOI:10.3152/147154405781776210)

Skyrms B, Pemantle R (2000) A dynamic model of social network formation. Proc Natl Acad Sci USA 97(16):9340–9346 (DOI: 10.1073/pnas.97.16.9340, stable JSTOR URL: http://www.jstor.org/stable/123177)

Snijders TAB (1990) Testing for change in a digraph at two time points. Soc Networks 12(4):359–373 (DOI: 10.1016/0378-8733(90)90015-2)

Snijders TAB (1995) Methods for longitudinal social network data. In: Tiit EM, Kollo T, Niemi H (eds) Multivariate statistics and matrices in statistics: Proceedings of the 5th Tartu Conference, Tartu, Pühajärve, Estonia, 23–28 May 1994. New Trends in Probability and Statistics, vol 3. TEV, Vilnius; VSP, Utrecht, pp 221–227

Snijders TAB (1996) Stochastic actor-oriented models for network change. J Math Sociol 21(1&2):149–176 (DOI: 10.1080/0022250X.1996.9990178)

Snijders TAB (2001) The statistical evaluation of social network dynamics. In: Sobel ME, Becker MP (eds) Sociological methodology: Volume 31: 2001. Blackwell, Boston, MA, pp 361–395 (DOI: 10.1111/0081-1750.00099, stable JSTOR URL: http://www.jstor.org/stable/3097281)

Snijders TAB, Steglich CEG, Schweinberger M, Huisman K (2008) Manual for SIENA version 3.2. University of Groningen, ICS, Department of Sociology, Groningen; University of Oxford, Department of Statistics, Oxford, available online at the URL: http://stat.gamma.rug.nl/s_man317.pdf and at the URL: http://www.stats.ox.ac.uk/ snijders/siena/s_man312.pdf

Snijders TAB, van de Bunt GG, Steglich CEG (2010) Introduction to stochastic actor-based models for network dynamics. Soc Networks 32(1):44–60 (DOI: 10.1016/j.socnet.2009.02.004)

Sonnenwald DH (2007) Scientific collaboration. Ann Rev Inf Sc Technol 41:643–681 (DOI: 10.1002/aris.2007.1440410121)

Tomassini M, Luthi L (2007) Empirical analysis of the evolution of a scientific collaboration network. Physica A 385(2):750–764 (DOI: 10.1016/j.physa.2007.07.028)

Torgerson WS (1952) Multidimensional scaling: I. Theory and method. Psychometrika 17(4):401–419 (DOI: 10.1007/BF02288916)

Travers J, Milgram S (1969) An experimental study of the small world problem. Sociometry 32(4):425–443 (DOI: 10.2307/2786545, stable JSTOR URL: http://www.jstor.org/stable/2786545)

Vázquez A (2003) Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. Phys Rev E 67(5):056104, 15 pp (DOI: 10.1103/PhysRevE.67.056104), also available as arXiv preprint arXiv:cond-mat/0211528

Veugelers R, Cassiman B (2005) R&D cooperation between firms and universities: Some empirical evidence from Belgian manufacturing. Int J Ind Organization 23(5–6):355–379 (DOI: 10.1016/j.ijindorg.2005.01.008), also available online as Katholieke Universiteit Leuven, Department Toegepaste Economische Wetenschappen Research Report 0325 at the URL: https://lirias.kuleuven.be/bitstream/123456789/118346/1/OR_0325.pdf

Wagner CS (2005) Six case studies of international collaboration in science. Scientometrics 62(1):3–26 (DOI: 10.1007/s11192-005-0001-0)

Wagner CS, Leydesdorff L (2005) Network structure, self-organization, and the growth of international collaboration in science. Res Policy 34(10):1608–1618 (DOI:10.1016/j.respol.2005.08.002), also available as arXiv preprint arXiv:0911.4299

Wang P, Robins G, Pattison P (2009) PNet: Program for the simulation and estimation of (p*) exponential random graph models: User Manual. University of Melbourne, Department of Psychology, Melbourne, available online at the URL: http://www.sna.unimelb.edu.au/pnet/pnet.html

Wasserman SS (1978) Models for binary directed graphs and their applications. Adv Appl Probab 10(4):803–818 (DOI: 10.2307/1426660, stable JSTOR URL: http://www.jstor.org/stable/1426660)

Wasserman S (1980a) Analyzing social networks as stochastic processes. J Am Stat Assoc 75(370):280–294 (DOI: 10.2307/2287447, stable JSTOR URL: http://www.jstor.org/stable/2287447)

Wasserman SS (1980b) A stochastic model for directed graphs with transition rates determined by reciprocity. In: Schuessler KF (ed) Sociological methodology 1980. Jossey-Bass, San Francisco, CA, pp 392–412 (DOI: 10.2307/270870, stable JSTOR URL: http://www.jstor.org/stable/270870).

Wasserman S (1987) Conformity of two sociometric relations. Psychometrika 52(1):3–18 (DOI:10.1007/BF02293953)

Wasserman S, Faust K (1994) Social network analysis: Methods and applications. Structural Analysis in the Social Sciences, vol 8. Cambridge University Press, Cambridge

Wasserman S, Iacobucci D (1988) Sequential social network data. Psychometrika 53(2):261–282 (DOI: 10.1007/BF02294137)

Wasserman S, Pattison P (1996) Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p. Psychometrika 61(3):401–425 (DOI:10.1007/BF02294547)

Wasserman S, Weaver SO (1985) Statistical analysis of binary relational data: Parameter estimation. J Math Psychol 29(4):406–427 (10.1016/0022-2496(85)90003-3)

Watts DJ (2003) Six degrees: The science of a connected age, 1st edn. W. W. Norton & Company, New York, NY

Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393(6684):440–442 (DOI: 10.1038/30918)

White D, Reitz K (1983) Graph and semigroup homomorphisms on networks of relations. Soc Networks 5(2):193–234 (DOI: 10.1016/0378-8733(83)90025-4)

Whitley R (2000) The intellectual and social organization of the sciences. Oxford University Press, Oxford

Wray KB (2002) The epistemic significance of collaborative research. Philos Sci 69(1):150–168 (DOI: 10.1086/338946, stable JSTOR URL: http://www.jstor.org/stable/3081018)

Xie F, Cui W (2008a) Cost range and the stable network structures. Soc Networks 30(1):100–101 (DOI: 10.1016/j.socnet.2007.07.002)

Xie F, Cui W (2008b) A note on the paper 'cost range and the stable network structures'. Soc Networks 30(1):102–103 (DOI: 10.1016/j.socnet.2007.07.005)

Ziman J (1994) Prometheus bound: Science in dynamic steady state. Cambridge University Press, Cambridge

Ziman J (2000) Real science: What it is, and what it means. Cambridge University Press, Cambridge (DOI: 10.1017/CBO9780511541391)

Zuckerman H (1967) Nobel laureates in science: Patterns of productivity, collaboration, and authorship. Am Sociol Rev 32(3):391–403 (DOI: 10.2307/2091086, stable JSTOR URL: http://www.jstor.org/stable/2091086)