Networks and Scientific Innovation

Thomas Heinze
Lehrstuhl für Organisationssoziologie,
Bergische Universität Wuppertal, Germany

Synonyms

Co-citation; Co-publication; Cumulative advantage; Homophily; Patents; Research productivity; Scientific elite; Trend-following

Introduction

The analysis of social networks in scientific innovation has seen a remarkable boom since the late 1990s: research on networks has developed into an interdisciplinary field comprising numerous mathematicians, physicists, and computer specialists, and no longer solely anthropologists, psychologists, and sociologists. A major reason for this boom is the availability of larger data sets and greater computer capacities for analyzing these data. Today, analyses quite commonly focus on co-publications with tens of thousands of researchers, co-citations between several million papers, or patent applications over periods of several decades. These data sets have improved the possibilities of investigating the mechanisms of network evolution and their role in scientific innovation (Chen and Redner 2010; Jones et al. 2008; Fleming et al. 2007; Wuchty et al. 2007; Newman et al. 2006; Powell et al. 2005).

This entry discusses key questions to approach selected findings from recent literature; a summary follows them: (1) What are networks in science and how are they defined? (2) What structures and characteristics do such networks have? (3) How do such networks arise and develop? (4) What is their role in scientific innovation?

What are Networks in Science?

In the terminology of mathematical graph theory, graphs consist of a finite number of nodes, connected by vertices. If all the vertices point in one direction, one speaks of a directed graph, otherwise, of an undirected graph. The number of vertices ending in a node is called the node degree; with directed networks, an indegree is distinguished from an outdegree. In the terminology of social network analysis, graphs are called networks, nodes are called actors, and vertices are called relationships.

When speaking of networks in science, one refers to collaboration among scientists, for example, in the framework of experiments, projects, or publications. Such cooperative relationships have the production and distribution of new knowledge or new technologies in the foreground. Here, the term *social networks* suggests itself. An indicator for social networks often used in the literature is joint authorship in the form of *co-publications* (*copub*). These are especially visible relationships that usually emerge from diverse formal and informal kinds of collaboration. Copub networks always consist of undirected relationships.

Networks in science also include intellectual connections among scientists arising through reference to the work results of colleagues. In the foreground of such reference relationships is usually the embedding of new arguments and findings within existing knowledge, where this is not based on collaborations. Here, the term *cognitive networks* suggests itself. An indicator for cognitive networks often used in the literature is *citation* (*cit*) *or co-citation* (*cocit*). Here, too, one deals with especially noted relationships to already published knowledge, which are far from being able to comprise all the real intellectual relations of a publication. Cit networks (A cites B, B cites C, etc.) always consist of directed relationships, while cocit networks are composed of undirected relationships (A and B cite C, B and C cite D, etc.).

Social and cognitive relationships can be analyzed not only on the microlevel of scientists. Empirical studies also investigate such relationships on higher levels of aggregation. These include research organizations, disciplinary communities, national research systems, and the global science system. The selection of the level of aggregation is generally determined by the knowledge the respective study is interested in. But analyses on a higher aggregation level also have the advantage of using temporal and disciplinary limitation to counter the long-familiar methodological problem of network analysis, namely, that there are no clearly derivable rules defining where a network begins and where it should end. With the *temporal limitation* to specific years or decades and the *factual limitation* to specific disciplinary communities (Chen and Redner 2010), research organizations (Jones et al. 2008; Heinze and Kuhlmann 2008), or – as in the case of the global science system – to selected databanks (Milojevic 2010; Jones et al. 2008), the boundaries of the networks to be investigated are defined pragmatically.

Structures and Characteristics of Networks in Science

Once the data basis is defined, the first important step of network analysis consists in investigating the fundamental structures and characteristics of the relationships. These include, in particular, the distribution of node degrees, the network's degree of differentiation, and its cohesion.

Distribution of Node Degree. For some years now, there has been intensive discussion about how the distribution of node degrees follows from copub and cit networks (Newman et al. 2006: 335ff). In many networks, the node degrees are not normally distributed around the mean, as they would be with the bell curve. Rather, there are many extreme values, so-called hubs, that is, actors who collaborate extremely often or articles that are extremely frequently cited. Networks with such hubs can be better described with the power law distribution (PLD), the distribution that is also valid for the productivity of scientists (Lotka's Law). But the PLD typically registers only observed values within a certain range of values that does not cover the entire distribution. In the case of a copub network in nanotechnology investigated by Milojevic (2010), for example, this value range lies between 20 and 200 coauthors. Below the threshold of 20 coauthors, there is a lognormal distribution.

The distribution of the node degrees is of great theoretical importance, because it is tied to the question of the mechanisms responsible for the rise and reproduction of network ties. There is a general consensus in the literature that the PLD results from the mechanism of *cumulative advantage* (CA), which was already described by Merton (1973). The hypothesis here is that scientists with higher node degrees are more likely to have new collaboration partners than are scientists with lower node degrees. Small initial differences grow over time into greater inequality. CA thus leads to a higher concentration of relationships in a few nodes. The close connection between PLD and CA means that whenever other distributions can be shown in addition to PLD, as in the case of Milojevic (2010), mechanisms other than CA are obviously at work in the genesis of the network. What mechanisms these are will be discussed below (cf. section "Mechanisms of Network Formation and Network Evolution").

Degree of Differentiation. There is also an intensive discussion about the effective identification of sub-communities and thematic fields within disciplines. In addition to the traditional procedures of social network analysis, for example, the analysis of cliques, clusters, or block models, in recent years a promising algorithm has been developed that identifies densely connected segments of the network without requiring knowledge of the content of the field covered by the network (Newman et al. 2006). Within a value area that is simple to interpret ($0 < Q < 1$), this *modularity* algorithm measures a network's degree of differentiation. For example, for the cit network of the journal family Physical Review, Chen and Redner (2010) calculate $Q = 0.543$, corresponding to 274 delimitable thematic areas. These thematic areas are in turn differentiated to different degrees. While high-temperature superconductivity ($Q = 0{,}198$) and Bose-Einstein condensation ($Q = 0.217$) have only a few subfields, metals/alloys ($Q = 0.481$) and quantum mechanics ($Q = 0.447$) are each markedly more differentiated.

Cohesion is another important concept for characterizing social and cognitive networks.

It is measured, on the one hand, by the average number of nodes lying between two randomly chosen nodes. As Newman (2001) is able to show for copub networks in various disciplines, the average distance is about six nodes and thus an order of magnitude comparable to that of other social, biological, and technological networks. In the global science system, a researcher thus needs only six intermediate steps to reach another, randomly selected researcher.

Another indicator for cohesion is the *cluster coefficient*, which measures the relative frequency of transitive triads (A publishes with B, B with D, and A with D). For the aforementioned copub networks (excepting in biology), Newman (2001) calculates probabilities between 30% and 70% that relationships A-B and B-D will result in a relationship A-D. These results are very similar to the idea of Granovetter (1973) that whenever strong relationships exist between A-B and B-D, there is social pressure on A-D to enter into a similarly directed relationship and thereby to bring about a *transitive triad* (also: closed triad). In the case that the relationship A-D does not come about, the social cohesion between A, B, and D is endangered. Granovetter (1973) coined the triple constellation that lacks the A-D relationship a *forbidden triad* (also: open triad) and points out that transitive triads arise only where relationships are strong. Where relationships between A-B and B-D are weak, A-D typically do not enter into a relationship; here, B remains a broker who mediates between A and D. Newman's results thereby indicate that, in the copub networks he investigated, between 30% and 70% of the relationships are strong. At the same time, Newman's findings indicate that here there is another mechanism leading to the rise of social relationships that effects the *formation of transitive triads* (FT). The extremely low probability of transitive triads in biology (7%) is an indication that in this discipline the majority of relationships are weak and biologists therefore do not customarily recruit new collaboration partners from the group of their own collaboration partners. Powell et al. (2005) confirm this finding (cf. section "Mechanisms of Network Formation and Network Evolution").

Mechanisms of Network Formation and Network Evolution

In the booming interdisciplinary context of network research on scientific innovation, an intensive discussion is being conducted on what mechanisms are crucial for the formation and evolution of networks. On this, the following discusses randomness, cumulative advantage, homophily, trend-following, and multiple connections.

Random Attachment. In many studies, randomly generated connections between actors play an important role. This is because mathematically oriented network analysis has always studied randomly generated graphs (model networks) and uses the characteristics it finds in them for comparisons with real networks (Newman et al. 2006: 229ff). But by far, not all the characteristics of randomly generated networks can be found in real networks of relationships. One especially striking deviation was found for the aforementioned cluster coefficients, where real networks often display a large multiple of what is measured in randomly generated networks. The reason for this deviation is the aforementioned FT mechanism, which ensures that real networks consist of many small clusters (cf. section "Structures and Characteristics of Networks in Science"). It is interesting that the high degree of cluster formation in real networks leads one to expect a relatively long average path length. This would mean that contacts spanning more than one cluster would be rare and that the actors would need long routes to reach an actor in another cluster. But as Watts (2003: 69ff) shows, the path lengths in real networks are typically quite short and differ only slightly from those in randomly generated networks. Many real networks, and especially copub networks, display high local densities and at the same time good global accessibility (Newman 2001). In the literature, networks with these two opposing characteristics are called "small worlds" (Newman et al. 2006: 9ff, 286ff).

How can relatively short path lengths arise despite the FT mechanism? Watts (2003: 83ff) argues that the short average path lengths could be produced by reconnecting existing relationships randomly. The underlying idea is simple: in networks with high local density, the probability that a random reconnection will produce a very distant relationship is quite high. This means that each reconnection very probably results in a connection with a previously unconnected cluster, which in turn reduces the average path length. The crux of the matter in this consideration is that randomness not only serves as heuristics for modeling the rise of real networks. Watts (2003) explicitly points out that forces of disorder and unforeseeability affect every real network, so that relationships among actors arise partially randomly. Taking this argument seriously, then, in the example of path length, randomness appears as a corrective to the FT mechanism. Thus, in the genesis of relationships between actors and in the dynamic of networks, random connections play a substantial role.

Cumulative Advantage. As already noted, the CA mechanism has the effect that already reputed and networked scientists can win new collaboration partners more frequently than less well-known or peripheral colleagues can. The logic of CA is thereby that small initial differences among researchers can grow over time to become a distribution in which a few researchers have a great many collaborative relationships and many colleagues have only a few (PLD). In the analysis of CA, progress has been made by carrying out a longitudinal study of extensive copub networks. For example, Barabási et al. (2002) analyze mathematics and the neurosciences on the global level in the years 1991–1998. New actors and relationships are added to the network each year, so Barabási et al. (2002) examine two sub-mechanisms. CA-1 means that young scientists co-publish with established researchers. Each increase of new researchers should thus lead to an increase in the average node degree. CA-2 means that the probability of a first-time collaboration between two established researchers within the network increases linearly with the frequency of their prior collaborations. CA-1 and CA-2 are both empirically confirmed.

Trend-Following and Homophily. In the literature, it is non-controversial that CA is an important element in the explanation of network emergence. However, in their study of the dynamics and evolution of inter-organizational networks between biotech companies in the period 1988–1999, Powell et al. (2005) identify additional social mechanisms. Trend-following (TF) means that one chooses the partner whom one's own circle perceives as attractive. Homophily (HP) means that partner selection is shaped by the principle that "birds of a feather flock together." Each of these mechanisms, however, has been only partially empirically confirmed. This means that, when choosing new partners, the biotech companies initially take their orientation from the conventions of their circles. Those partners are selected whom the circle perceives as attractive. But TF is not valid for repeated contacts; here, the biotech companies manage to emancipate themselves from the trend. A similar pattern emerges for HP. New contacts are extremely frequently begun with spatially close partners, but spatial closeness plays no role for repeated contacts.

Multiple Connections. Whether a biotech company repeats its collaboration with a partner depends, rather, on whether the partner brings diversity into the relationship and whether the partnership holds promise of long-term gains. Multiple connections thus mean, first, a preference for heterogeneity in choice of partners (MC-1) and, second, a preference for deepening existing partnerships (MC-2). As Powell et al. (2005) show, in biotechnology or the life sciences, there is a marked preference for competences and contact structures that one does not possess oneself. Collaboration partners with a diverse contact portfolio are thus especially attractive, because they open up access to new knowledge and new technologies. The great preference for heterogeneous knowledge and know-how is reflected in the fact that young beginners are especially coveted, in contrast to established biotech companies (MC-1). However, Powell et al. (2005) also show that, once a high level of diversity is achieved, the search for new partners slackens. In this case, the biotech company deepens its relationships to its partners and bonds them to it for the long term (MC-2). As a social mechanism that steers the formation and continuation of relationships in networks, multiple connections thus entail a tension between the search for new knowledge and know-how, on the one hand, and the search for a stable and fruitful partnership, on the other. Overall, the results of Powell et al. (2005) indicate that MC-1 and MC-2, rather than CA, are the dominant social mechanisms that explain the rise and evolution of inter-organizational partnerships in the biotech sector. The authors thereby confirm Newman's (2001) finding that, in biology and the life sciences, it is less customary than in other disciplines to make contacts within the circle of one's own collaboration partners (cf. section "Structures and Characteristics of Networks in Science").

Networks and Scientific Innovation

Networks are not only important in the biotech sector but also in other scientific fields and disciplines. For example, Fleming et al. (2007) examine the collaborative networks of inventors in the United States, based on 2.8 million patent specifications from the years 1975 to 2002. The starting point for this study is the question whether brokered structures with open triads or cohesive structures with closed triads increase the productive capacity of networks (cf. section "Structures and Characteristics of Networks in Science"). The authors show that collaborative networks with brokers often lead to technical innovations. At the same time, however, technical innovations from brokered networks are less frequently used again than are those from cohesive networks. These results indicate that new knowledge spreads better in socially integrated contexts, while brokered contexts create hurdles for the spread of new ideas. Fleming et al. (2007) point out that there is a paradox here, namely, that the network structures suitable for developing technical innovations are not suitable for their diffusion, while vice versa those network structures that are unsuitable for bringing about technical innovations are especially suitable for spreading them. Fleming et al. (2007) sketch a possible escape from this paradox: recruiting actors in cohesive networks who have a broad spectrum of knowledge, have gathered experience in various organizations, and also initiate contacts outside their own work contexts. In this way, the structural disadvantages of cohesive networks in giving rise to new ideas can be at least partially compensated.

Networks also influence scientific productivity capacity and rankings in research. Jones et al. (2008) show in their analysis of the 662 largest universities in the United States that, based on the Web of Science, in the period 1975–2005 interuniversity copub relationships more than doubled, both among the natural and engineering sciences and among the social sciences. Today, about a third of all papers are published by interuniversity teams. This growth derives essentially from decades of the generally increasing proportion of co-publications in the global science system. Also based on the Web of Science, Wuchty et al. (2007) calculate that, in the period 1955–2000, the number of co-publications in the social sciences rose from 18% to 52% and in the natural and engineering sciences from 50% to 83%. At the same time, the average number of coauthors in the social sciences increased from 1.3 to 2.3 and in the natural and engineering sciences from 1.9 to 3.5.

As Jones et al. (2008) further show, interuniversity publications are cited substantially more often than are publications by authors who all belong to a single university. The greater visibility of interuniversity publications is unequally distributed: the greater the number of citations from a site, the more it profits from interuniversity collaborations. This means that the effect of interuniversity publications on visibility and thus also on scientific prestige is concentrated on elite organizations. Here, the gap in visibility and prestige between elite and periphery increased markedly in the period 1975–2005. The increasing density of interuniversity copub networks thus amplifies the already marked institutional stratification of the university system in the United States. Finally, Jones et al. (2008) show that collaborations between different sites of the university elite (and incidentally also, separately, between peripheral sites) are more frequent than mixed relationships. This indicates that the aforementioned HP mechanism decisively shapes the genesis of interuniversity relationships (cf. section "Mechanisms of Network Formation and Network Evolution").

Conclusion and Future Directions

In sum, it can be noted that many interesting things about network formation, network evolution, network structures, and their influence on innovative science are known. The availability of large longitudinal data sets makes it possible to conceptualize and empirically examine the connection between the statistical distribution of cognitive and social relationships, the mechanisms of their emergence and reproduction, and their role in fostering research productivity and scientific innovation. The wide spectrum of investigated networks has also resulted in a better understanding of the cultural differences between various disciplines and fields of research. Good examples of this are the life sciences, whose networks markedly differ from other disciplines, in particular from physics (Powell et al. 2005; Newman 2001).

With regard to the aggregation levels mentioned at the beginning of this entry, recent research on networks has produced some studies of the global science system (Wuchty et al. 2007; Newman 2001), but the majority of the analyses still focus on disciplines and fields of research (Chen and Redner 2010; Milojevic 2010). In this regard, recent interdisciplinary research on networks follows an established path to the disadvantage of research organizations. There are only a few studies that address the theme of universities or non-university institutes, including industry research, as nodes of social or cognitive networks and their role in scientific innovation (Jones et al. 2008; Powell et al. 2005). There is a clear need to pay more attention to the organizational level with regard to networks and scientific innovation in the future.

Cross-References

▶ Adaptive Creativity and Innovative Creativity
▶ Creativity and Innovation: What Is the Difference?
▶ Innovation Policies (vis-à-vis Practice and Theory)
▶ Innovation Policy Learning
▶ Innovations of and in Organizations
▶ Research on Creativity
▶ Science of Creativity

References

Barabási A-L, Jeong H, Néda Z, Ravasz E, Schubert A, Vicsek T. Evolution of the social network of scientific collaborations. Physica A. 2002;311:590–614.
Chen P, Redner S. Community structure of the physical review citation network. J Informetr. 2010;4:278–90.
Fleming L, Mingo S, Chen D. Collaborative brokerage, generative creativity, and creative scientists. Adm Sci Q. 2007;52:443–75.
Granovetter M. The strength of weak ties. Am J Sociol. 1973;78:1360–80.
Heinze T, Kuhlmann S. Across institutional boundaries? Research collaboration in German public sector nanoscience. Res Policy. 2008;37:888–99.
Jones BF, Wuchty S, Uzzi B. Multi-university research teams: shifting impact, geography, and stratification in science. Science. 2008;322:1259–62.
Merton RK. The normative structure of science. In: Storer NW, editor. The sociology of science: theoretical and empirical investigations. Chicago: The University of Chicago Press; 1973. p. 267–78.
Milojevic S. Modes of collaboration in modern science. Beyond power laws and preferential attachment. J Am Soc Inf Sci Technol. 2010;61:1410–23.
Newman M. The structure of scientific collaboration networks. Proc Natl Acad Sci. 2001;98:404–9.
Newman M, Barabási A-L, Watts D. The structure and dynamics of networks. Princeton: Princeton University Press; 2006.
Powell WW, Koput KW, White DR, Owen-Smith J. Network dynamics and field evolution. The growth of interorganizational collaboration in the life sciences. Am J Sociol. 2005;110:1132–205.
Watts D. Six degrees. The science of the connected age. New York: Norton & Norton; 2003.
Wuchty S, Jones BF, Uzzi B. The increasing dominance of teams in production of knowledge. Science. 2007;316:1036–9.

Networks Interfirms Cooperation

▶ Partnerships and Entrepreneurship (Vol Entrepreneurship)