Mehmet Çakırtaş
Mehmet Kemal Özdemir   *Editors*

Big Data and Social Media Analytics

Trending Applications

Springer

Lecture Notes in Social Networks

Lecture Notes in Social Networks (LNSN) comprises volumes covering the theory, foundations and applications of the new emerging multidisciplinary field of social networks analysis and mining. LNSN publishes peer-reviewed works (including monographs, edited works) in the analytical, technical as well as the organizational side of social computing, social networks, network sciences, graph theory, sociology, Semantics Web, Web applications and analytics, information networks, theoretical physics, modeling, security, crisis and risk management, and other related disciplines. The volumes are guest-edited by experts in a specific domain. This series is indexed by DBLP. Springer and the Series Editors welcome book ideas from authors. Potential authors who wish to submit a book proposal should contact Christoph Baumann, Publishing Editor, Springer e-mail: Christoph.Baumann@springer.com

More information about this series at http://www.springer.com/series/8768

Mehmet Çakırtaş • Mehmet Kemal Ozdemir
Editors

Big Data and Social Media Analytics

Trending Applications

Springer

Editors
Mehmet Çakırtaş
Bilkent yerleşkesi
Turkish Ministry of Health
Çankaya, Ankara, Turkey

Mehmet Kemal Ozdemir
Computer Engineering
Istanbul Medipol University
Istanbul, Turkey

Contents

Twenty Years of Network Science: A Bibliographic and Co-authorship Network Analysis

Roland Molontay ⬤ **and Marcell Nagy** ⬤

Abstract Two decades ago three pioneering papers turned the attention to complex networks and initiated a new era of research, establishing an interdisciplinary field called network science. Namely, these highly-cited seminal papers were written by Watts and Strogatz, Barabási and Albert, and Girvan and Newman on small-world networks, on scale-free networks and on the community structure of complex networks, respectively. In the past 20 years – due to the multidisciplinary nature of the field – a diverse but not divided network science community has emerged. In this chapter, we investigate how this community has evolved over time with respect to speed, diversity and interdisciplinary nature as seen through the growing co-authorship network of network scientists (here the notion refers to a scholar with at least one paper citing at least one of the three aforementioned milestone papers). After providing a bibliographic analysis of 31,763 network science papers, we construct the co-authorship network of 56,646 network scientists and we analyze its topology and dynamics. We shed light on the collaboration patterns of the last 20 years of network science by investigating numerous structural properties of the co-authorship network and by using enhanced data visualization techniques. We also identify the most central authors, the largest communities, investigate the spatiotemporal changes, and compare the properties of the network to scientometric indicators.

Keywords Science of science · Network science · Bibliometrics · Scholarly data · Scholarly network analysis · Co-authorship network

R. Molontay (✉)
Department of Stochastics, Budapest University of Technology and Economics, Budapest, Hungary

MTA–BME Stochastics Research Group, Budapest, Hungary
e-mail: molontay@math.bme.hu

M. Nagy
Department of Stochastics, Budapest University of Technology and Economics, Budapest, Hungary
e-mail: marcessz@math.bme.hu

1

1   Introduction

Complex networks have been studied extensively since they efficiently describe
a wide range of systems, spanning many different disciplines, such as Biology
(e.g. protein interaction networks), Information Technology (e.g., WWW, Internet),
Social Sciences (e.g., collaboration, communication, economic, and political net-
works), etc. Moreover, not only the networks originate from different domains,
but the methodologies of network science as well, for instance, it heavily relies
on the theories and methods of graph theory, statistical physics, computer science,
statistics, and sociology.

In the last two decades, network science has become a new discipline of great
importance. It can be regarded as a new academic field since 2005 when the
U.S. National Research Council defined network science as a new field of basic
research [11]. The most distinguished academic publishing companies announce the
launch of new journals devoted to complex networks, one after another (e.g. Journal
of Complex Networks by Oxford University Press, Network Science by Cambridge
University Press, Applied Network Science and Social Network Analysis and
Mining by Springer). Network science also has its own prestigious conferences
attended by thousands of scientists. Leading universities continuously establish
research centers and new departments for network science, furthermore, launch
Master and Ph.D. programs in this field (such as Yale University, Duke University,
Northeastern University, and Central European University).

The significance of network theory is also reflected in the large number of
publications on complex networks and in the enormous number of citations of
the pioneering papers by Barabási and Albert [5], Watts and Strogatz [35] and
Girvan and Newman [13]. Some researchers interpret network science as a new
paradigm shift [16]. However, complex networks are not only acknowledged by the
research community, but innovative textbooks aimed for a wider audience have also
been published [7, 23], moreover, the concepts of network science have appeared in
the popular literature [4, 34] and mass media [30] as well.

In the last two decades, complex networks became in the center of research
interest thanks to – among many others – the aforementioned three pioneering
papers and due to the fact that the prompt evolution of information technology has
opened up new approaches to the investigation of large networks. This period of
20 years can be regarded as the golden age of network science. The first challenge
was to understand network topology, to this end, structural properties were put under
the microscope one after the other (small-worldness, scale-free property, modularity,
fractality, etc.) and various network models were proposed to understand and to
mathematically describe the architecture and evolution of real-world networks [33].
In recent years, there has been a shift from the structural analysis to studying the
control principles of complex networks [3]. Remarkable computing power, massive
datasets, and novel computational techniques keep great potential for network
scientists for yet another 20 years [33].

This work is a tribute to the achievements of the network science community in
the past 20 years. We provide a bibliographic analysis of 31,763 network science

papers and we also construct and investigate the co-authorship network of network scientists to identify how the network science community has been evolving over time.

The present study also extends the earlier conference version of this chapter [21] in several important directions. Namely, here we provide a more detailed literature review; we examine a longer time period; and we answer the question of how the network science community has evolved over time with respect to speed, diversity, and interdisciplinary nature by implementing novel analyses. Here we also provide an analysis of the co-occurrence network of the keywords. Moreover, we supplement our previous work with several other new methods and data visualizations that help to make insightful observations regarding the last two decades of network science.

The main contributions of this work can be summarized as follows:

– We collect 31,763 network science papers and provide a bibliographic analysis investigating various characteristics of the papers and showing how the discipline has developed over time.
– We construct the co-authorship network of 56,646 network scientists and undertake a scholarly network analysis study by analyzing its topology and dynamics.
– We answer the following major research questions:

  • What are the most important venues of network science and how have they changed over time?
  • How the publication patterns vary over research areas and time?
  • What are the most important topics of network science and how have they evolved through time? What relationships can we explore among the most frequent keywords of network science?
  • How the network science community has evolved over time with respect to speed, diversity, and interdisciplinary nature?
  • What are the most typical patterns in terms of international and interdisciplinary collaborations?
  • Who are the most central authors and how do the largest communities look like? How do these network properties compare to other scientometric indicators?

2   Scholarly Networks Analysis

The present paper joins the line of research focused on scholarly network analysis that is based on big scholarly data [28, 37]. Big scholarly data refers to the rapidly growing data accessible in digital libraries containing information on millions of scientific publications and authors [36]. The easily available data sources (Web of Science, Scopus, PubMed, Google Scholar, Microsoft Academic, the U.S. Patent and Trademark Office, etc.) together with novel powerful data analysis technologies have led to the emergence of science of science [12] that gives us a better understanding of the self-organizing rules and patterns of science, e.g. how

disciplines have emerged and evolved over time [19]. Various scholarly networks at many levels can be formed based on scholarly data; Pawar et al. identify the following forms of scholarly networks of great interest [28]:

1. co-authorship networks (a link is formed between scientists by their co-authorship of at least one scientific paper),
2. citation networks (a directed link is formed between documents referencing one another),
3. co-citation networks (a link is formed between documents if they are cited together),
4. bibliographic coupling (documents are linked if they share common references),
5. co-occurrence networks (keywords/topics are linked if they occur in the same document), and
6. heterogeneous networks (two or more coupled scholarly networks).

Among the aforementioned scholarly networks perhaps co-authorship networks have attracted the greatest deal of research interest, owing to the fact that co-authorship is one of the most important reflections of research collaboration, which is an essential mechanism that joins together distributed knowledge and expertise into novel discoveries. Furthermore, building the map of sciences is not only important for sociologists and other scholars to understand researchers' interaction but for policymakers as well to address sharing resources [36]. Co-authorship networks have been studied extensively in various ways and from various perspectives: e.g. the collaboration network determined by the articles of a certain journal, a specific country or a research community that cites a particular influential paper or author [1, 2, 6, 17, 24, 25]. In this chapter, we investigate the co-authorship network of network scientists as defined in the following section.

Keywords co-occurrence networks have also been investigated thoroughly [15, 20, 29, 31]. Keywords of academic articles can provide a concise overview of the content and the core idea of the body of the papers. In contrast to word clouds, co-occurrence networks do not only show the frequency of the keywords but also allow us to discover the relationship between them. Li et al. investigated 6,900 articles published between 1982 and 2013 which had been indexed using the keyword 'complex network(s)' and provided a co-keyword network and a keyword co-occurrence network analysis [20].

3   Preliminaries and Data

In this section, we describe how the co-authorship network of network scientists was constructed, how the examined set of academic publications was chosen and collected, what data preparation steps were conducted. Moreover, we also present some useful notions and preliminaries.

3.1 Co-authorship Network of Network Scientists

To the best of our knowledge, the co-authorship network of network scientists has been analyzed only by Newman et al. [26, 27]. However, their network consists of 1,589 authors, while this study investigates a much larger network (56,646 vertices and 357,585 edges) spanning a longer time horizon (1998–2019).

We construct the co-authorship network of network scientists as follows. We consider three ground-breaking papers around the millennium that can be regarded as the roots of the rise of network science: the paper of Watts and Strogatz [35] on small-world networks, the work of Barabási and Albert [5] about scale-free networks and the paper of Girvan and Newman [13] that reveals the community structure of complex networks. We selected the aforementioned three papers since they initiated new areas of research in network science by introducing pivotal concepts two decades ago, that had a huge impact on the network science community that is also demonstrated by the large number of citations they received in the past 20 years.

In this work, we consider a paper as a *network science paper* if it cites at least one of the three aforementioned pivotal articles (in addition, the three originating papers are also regarded as network science papers, obviously). Similarly, we call someone a *network scientist* if (s)he has at least one network science paper. The previous definitions of a network science paper and a network scientist are of course quite arbitrary. It is important to note that the papers that refer to one of the three seminal papers are not necessarily about network science and there certainly exist network science articles that do not refer to any of the aforementioned pioneering papers. On the other hand, we believe that this concept is a good proxy for our purposes and it is worth studying. We construct the co-authorship network of the network scientists where two of them are connected if they have at least one joint network science paper (see Fig. 1). In other words, this network is a one-mode projection onto scientists, from the bipartite network of scientists and the network science papers they authored. The anonymized data of the constructed network and some figures in high resolution are available in the supplementary material [22].

3.2 Glossary

Definition 1 (Graph) A simple (undirected) graph is an ordered pair $G = (V, E)$, where V is the set of vertices or nodes and E is the set of edges or links, which are two-element subsets of V. The vertex and edge sets of G are denoted by $V(G)$ and $E(G)$ respectively. The size of the graph is the number of its nodes, and it is usually denoted by n.

Definition 2 (Complex network) In network theory, the terms graph and network are used interchangeably, however, a complex network is a graph with non-trivial topological features, that characterize real-world networks.

Fig. 1 The co-authorship
network of network scientists
colored by communities
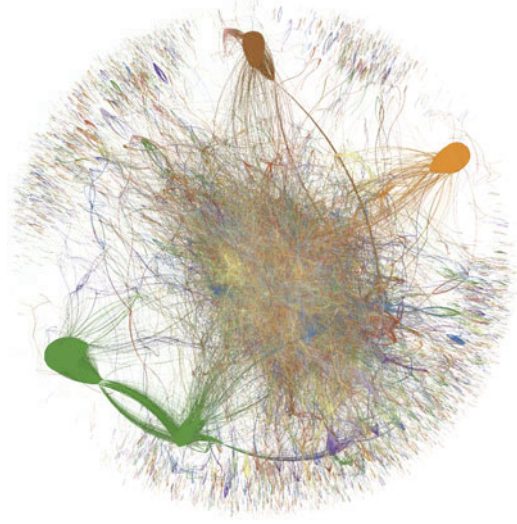
Definition 3 (Average path length) A path is a sequence of edges which connect a
sequence of vertices. The distance $d(u, v)$ between the vertices u and v is the length
(number of edges) of the shortest path connecting them. The l_G average path length
of a graph G of size n is defined as:

$$l_G = \frac{1}{n(n-1)} \sum_{\substack{u,v \in V(G) \\ u \neq v}} d(u, v).$$

Definition 4 (Small-world property) A network is said to be small-world, if the
average path length is proportional to the logarithm of the size of the network i.e.
$l_G \sim \log |V|$.

Definition 5 (Degree distribution) The degree $\deg(v)$ of a vertex v in a simple,
undirected graph is its number of incident edges. The degree distribution P is the
probability distribution of the degrees over the whole network, i.e. $P(k)$ is the
probability that the degree of a randomly chosen vertex is equal to k.

Definition 6 (Scale-free property) A scale-free network is a connected graph
which $P(k)$ degree distribution follows a power law asymptotically, i.e. $P(k) \sim
k^{-\gamma}$, where $\gamma \geq 1$.

Definition 7 (Assortativity coefficient) The assortativity coefficient is the Pearson
correlation coefficient of degree between pairs of linked nodes. The assortativity
coefficient is given by

$$r = \frac{\sum_{j,k} j \cdot k(e_{j,k} - q_j q_k)}{\sigma_q^2},$$

where the term q_k is the mass function of the distribution of the remaining degrees (degree of the nodes minus one) and j and k indicates the remaining degrees. Furthermore, $e_{j,k}$ refers to the mass function of the joint probability distribution of the remaining degrees of the two vertices. Finally, σ_q^2 denotes the variance of the remaining degree distribution with mass function q_k i.e. $\sigma_q^2 = \sum_k k^2 q_k - \left(\sum_k k q_k\right)^2$.

Definition 8 (Local clustering coefficient) The local clustering coefficient of vertex v is the fraction of pairs of neighbors of v that are connected over all pairs of neighbors of v. Formally:

$$C_{\text{loc}}(v) = \frac{|\{(s, t) \text{ edges} : s, t \in N_v \text{ and } (s, t) \in E|\}}{\deg(v)(\deg(v) - 1)},$$

where N_v is the neighborhood of the node v i.e. the vertices adjacent to v.

The average (local) clustering coefficient of a G graph is defined as:

$$\bar{C}(G) = \frac{1}{n} \sum_{v \in V(G)} C_{\text{loc}}(v),$$

where n is the size of the graph.

Definition 9 (Global clustering coefficient) The global clustering coefficient C of the graph G is the fraction of closed triplets (paths of length two in G that are closed) over all of the triplets (paths of length two) in G.

Definition 10 (Betweenness centrality) Betweenness centrality of a node v is given by the expression:

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}},$$

where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths that pass through v.

Definition 11 (h-index) The h-index is an author-level metric defined as the maximum value of h such that the given author has published h papers that have each been cited at least h times.

Definition 12 (Harmonic centrality) Harmonic centrality of a node v is the sum of the reciprocal of the shortest path distances from all other nodes to v:

$$H(v) = \sum_{u \neq v} \frac{1}{d(u, v)}.$$

3.3 Data Collection and Preparation

We build our analysis on data collected from the Web of Science bibliographic database, retrieved on January 2, 2020. The collected data consist of 41,245 rows with multiplicity corresponding to the citing works of the three seminal articles [5, 13, 35]. For each citing paper we have information on the document title, publication name, publication type, publisher, publication year, authors' full name, authors' address, research area, keywords, cited reference count, total times cited count, page count, abstract, etc.

After the data were collected, various data preparation steps were conducted, including merging the files, handling missing fields, deleting duplicates, and indicating which of the three seminal papers were cited by the given article. These preparation steps reduced the dataset to 31,763 unique rows and the citation pattern of the corresponding articles is shown in Fig. 2.

The authors are represented by the full name field of Web of Science, however, this field is unfortunately not consistent, the author called John Michael Doe may appear as Doe, John; John Doe; Doe, J.; Doe, J. M.; Doe, John Michael, and other variants. To overcome this issue, we created a dictionary that defines the name variants that correspond to the same author. Furthermore, we cannot distinguish between different scientists with the same name, this issue is mainly relevant for Asian authors. However, the error introduced by this problem is negligible, as also pointed out by Newman [24] and by Barabási et al. [6].

Fig. 2 Distribution of the citations among the three pioneering papers
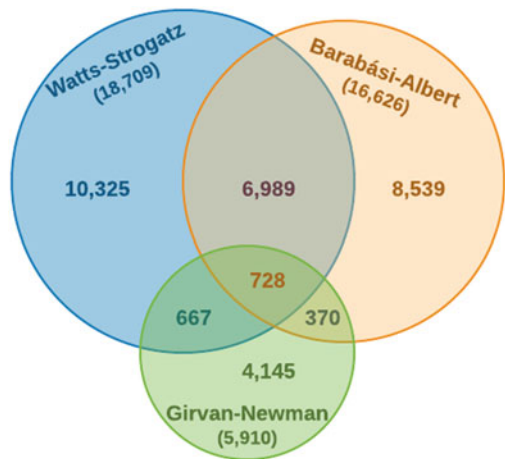
4 Analysis of Network Science Papers

First, we analyze the enormous number of citing works, i.e. the network science papers. Figure 3 shows the top 10 research areas where the citing works belong to, illustrating the interdisciplinary nature of network science. We can see that the first decade was dominated by physics while later computer science took over. It is also clear from the figure that neuroscience has started to use tools of network science in the last decade. The journals that publish the most network science papers are shown in Fig. 4. Considering the number of publications, Physical Review E was the leading scientific forum of network science in the first half of the examined period, while PLOS One and Scientific Reports emerged in the last decade. Currently, Physica A can be regarded as the leading journal of network science in terms of the number of published network science articles.

Figure 5 shows the number of collaborating authors per citing works, the most typical numbers of co-authors in a network science paper are 2 and 3. Almost one-tenth of the network science papers are sole-authored, M. E. J. Newman has the highest number of sole-authored network science papers, namely 27. While the figure shows only up to 15 number of authors, there are a few papers with a high number of collaborating authors e.g. the paper with the highest number of authors (387) is a paper of the Alzheimer's Disease Neuroimaging Initiative [18]. The authors of this article emerge as a maximal clique of the co-authorship network of network scientists as it can be seen in Fig. 1.

We also investigate the distribution of network science papers written by network scientists. The authors with the highest number of network science papers together with the citation count of their network science papers are shown in Table 1.
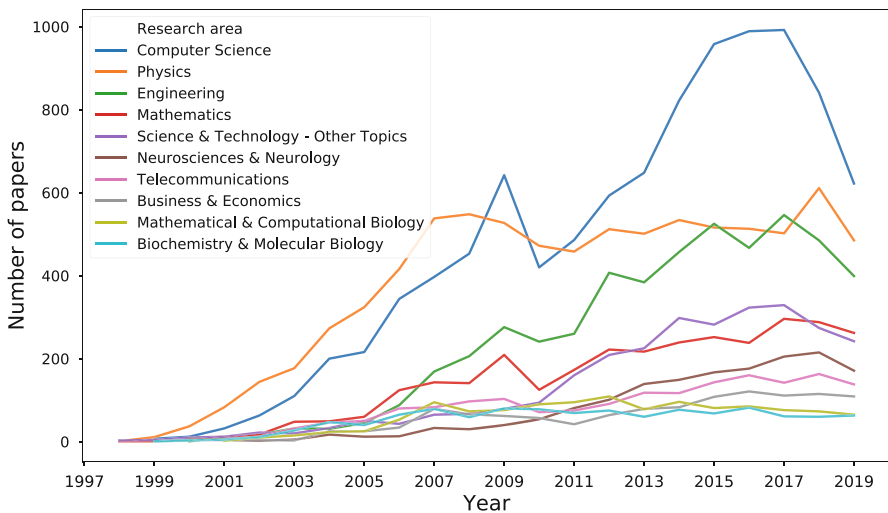

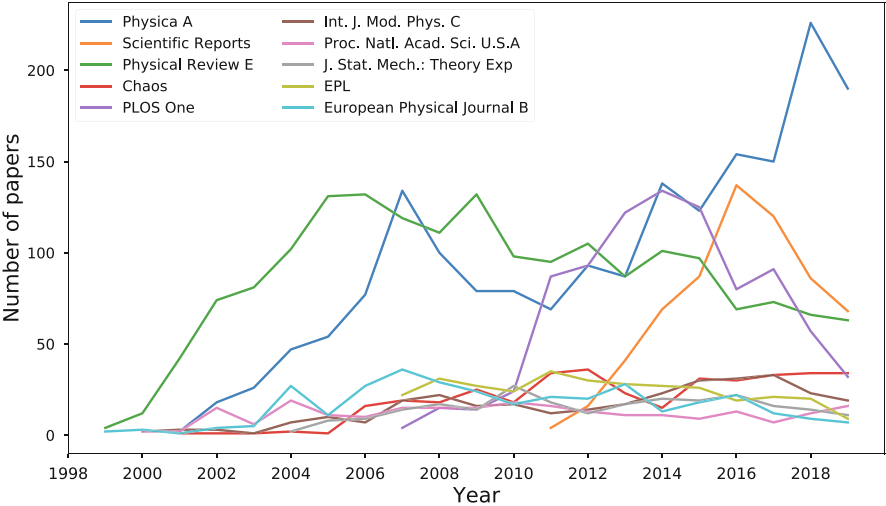**Fig. 3** Top 10 research areas of the network science papers

Fig. 4 Top 10 journals of the network science papers
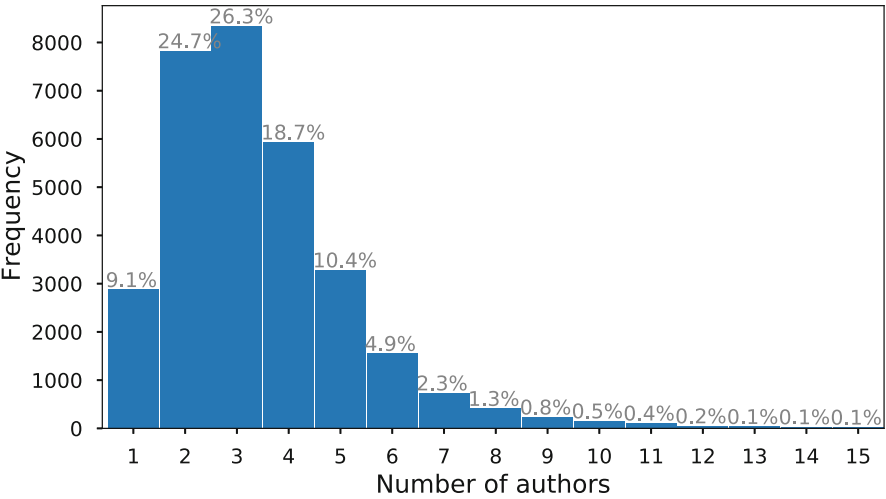
Fig. 5 Histogram of the number of authors per paper (truncated at 15)

Guanrong Chen has the highest number of network science papers, his research areas are nonlinear systems and complex network dynamics and control.

To gain some insights on how the publication patterns vary in network science depending on the research area, we show how the distribution of the number of cited references and the length of the papers differ across research areas (see Fig. 6). We can observe that in neuroscience and neurology authors typically cite a high number of articles while in computer science or engineering the typical number of

Table 1 Top 12 authors with the most network science papers

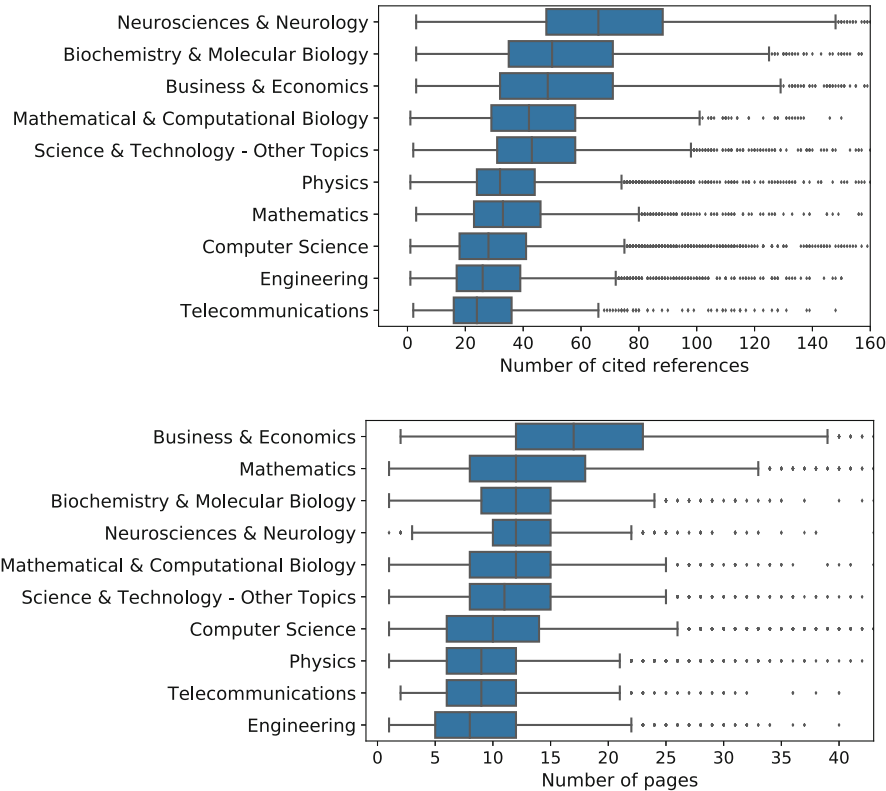| Name of author	Number of papers	Number of citations
Guanrong Chen	167	12,859
Bing-Hong Wang	145	5,341
Tao Zhou	138	9,911
Shlomo Havlin	124	13,377
Jürgen Kurths	118	9,249
Eugene H. Stanley	113	10,479
Zhongzhi Zhang	104	2,099
Ying-Cheng Lai	99	5,799
Albert-László Barabási	95	73,937
Luciano da Fontoura Costa	92	2,726
Matjaz Perc	89	8,634
Michael Small	88	2,345

Fig. 6 Boxplots of the number of cited references and length of network science papers across research areas
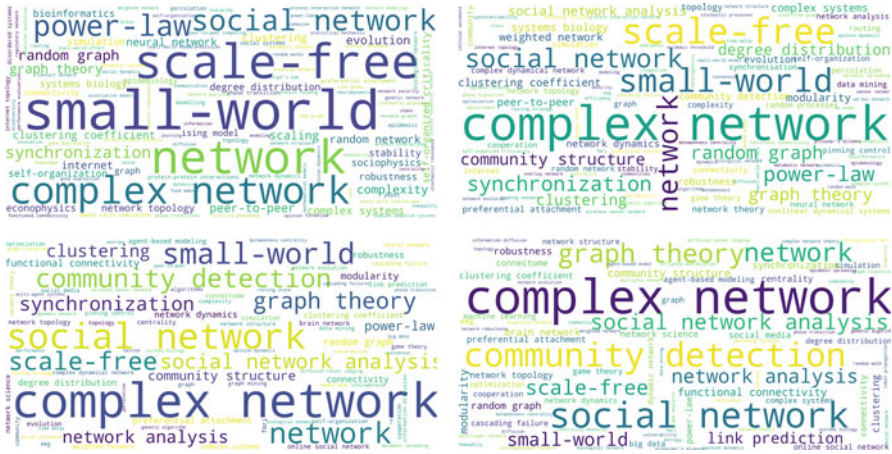
Fig. 7 Word cloud of the most frequent keywords of network science papers before 2006 (top right), between 2006 and 2010 (top left), between 2011 and 2015 (bottom left) and since 2016 (bottom right)

cited references is much smaller. Engineering together with telecommunications and physics are also the areas with the shortest articles, while it can also be due to the fact that in those disciplines double-column publication formats are quite typical. It is important to emphasize that these observations are not necessarily representative of the disciplines in general, only for those papers that were defined as network science papers.

Figure 7 depicts separate word clouds of the most frequently used keywords of network science papers written in the four 5-year-long periods of the last two decades of network science. We can observe that in the first half of the examined period structure related (e.g., scale-free, small-world, topology) and modeling related keywords (e.g., preferential attachment, evolution, growing network, small-world model) dominated the study of complex networks, while in the last decade topics such as community detection, social network analysis, data-driven research (big data, data mining, link prediction, machine learning) have become more popular keywords. In the first decade of the examined period, the most studied real-world networks were Internet, peer-to-peer, and protein interaction networks, however, since 2010 the research has tended to focus on online social networks and brain networks. These word clouds can also be found in the supplementary material [22].

To gain a more comprehensive understanding of the keywords of network science papers, we also construct their co-occurrence network. Figures 8 and 9 together with Fig. 7 clearly shows that 'complex network' is the most important keyword of the examined set of papers that supports our hypothesis that our definition of network science paper was indeed a good proxy of our purposes. We can observe that the keyword 'random graph' often goes together with 'network' but less often with
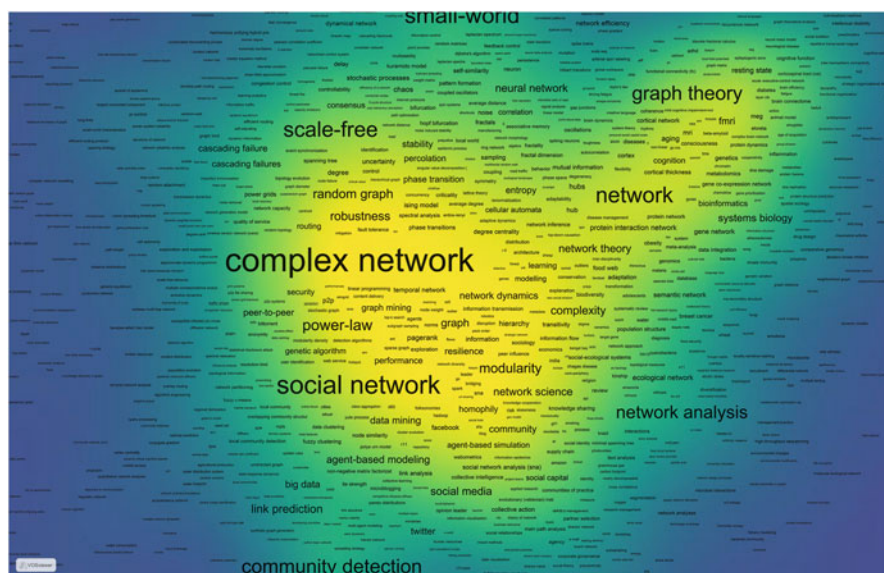
Fig. 8 Visualization of two-dimensional knowledge map of keywords. Keywords that have co-occurred more frequently are placed closer to each other on the map. The font size indicates the number and strength of the connections of a keyword. A more intense color implies a larger number of keywords and higher connectivity in the neighborhood of the point. The figure was created with VOSviewer [32]
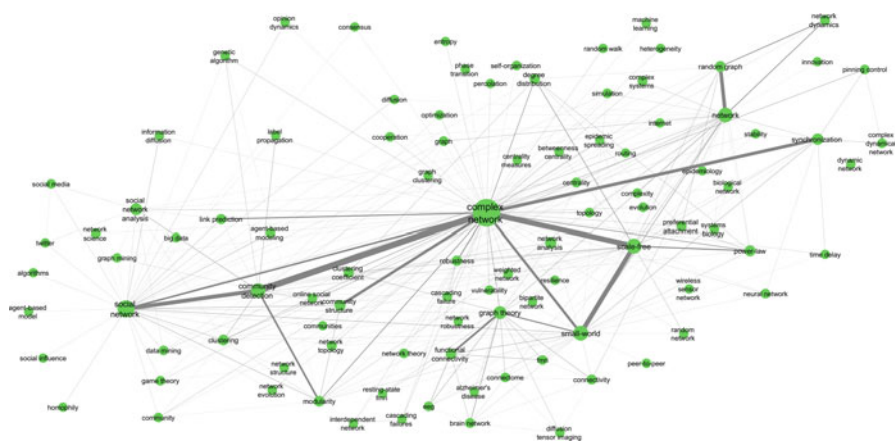

Fig. 9 Co-occurrence network of keywords. The size of the node indicates the frequency of keywords in network science papers, the edge width indicates their relative co-occurrence. Only keywords with frequency at least 100 and edges with weight at least 10 are shown in the figure
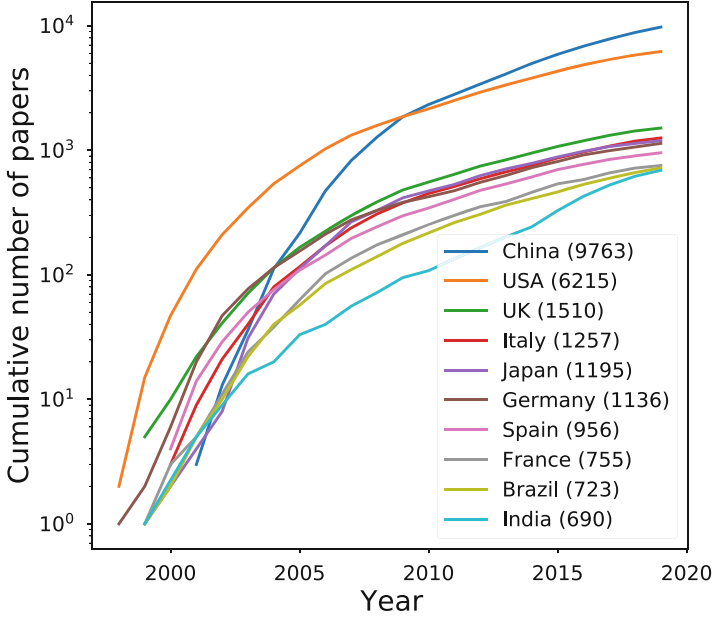
Fig. 10 Cumulative number of network science papers on a logarithmic scale by the country of the first author (only the Top 10 countries are shown)

'complex network' that suggests the term 'complex network' is not that widespread among mathematicians. Another observation is that community detection is rather popular in the social network domain, the bottom left side of the figures is dominated by terms associated with social network analysis and community detection. We can also observe that the keywords 'scale-free' and 'small-world' are frequently used together. The complete keyword co-occurrence network and the Figs. 8 and 9 can be found in the supplementary material [22].

Based on the address of the first author, we identify the network science hot-spots and investigate the spatiotemporal changes. Figure 10 demonstrates that China and the USA are the two leading countries of network science with a fast increase in Chinese network science papers in the last few years.

Figure 11 illustrates how the ratio of multidisciplinary and international papers varied over the years. We can observe that network science research gets both increasingly international and multidisciplinary. Here we consider a paper international if it has at least two co-authors who do not share an affiliation within the same country. The ratio of international papers in an important indicator, since it was also shown that scientific impact increases if researchers publish in international collaboration [8]. The multidisciplinary nature of papers is defined by the fact that more than one research area is attached to the document in Web of Science.
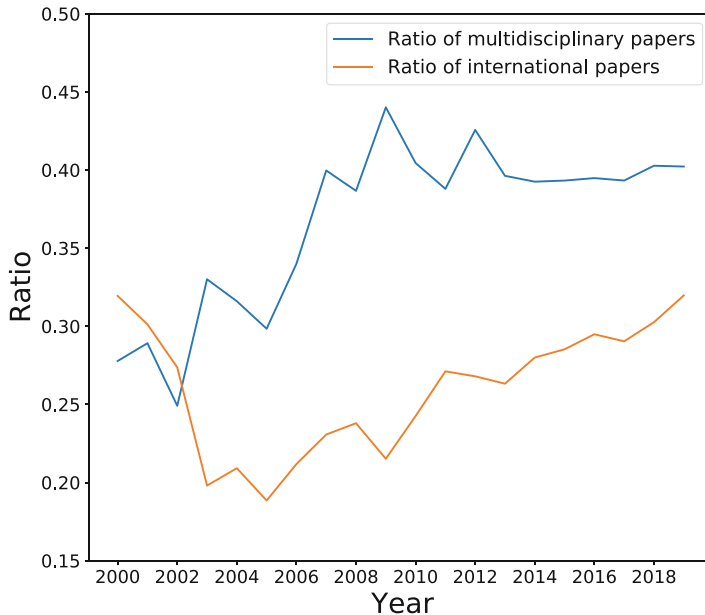
Fig. 11 Ratio of multidisciplinary and international papers since 2000

5 Analysis of the Co-authorship Network

The nodes of the co-authorship network of network scientists correspond to the authors who have at least one network science paper (i.e., a paper that cites at least one of the three seminal papers [5, 13, 35]), two of them are connected if they co-authored at least one network science paper. The network is simple, undirected, and unweighted meaning that here we ignore the strength of the connection between two scientists, i.e. the number of their joint papers. The network has 56,646 nodes and 357,585 edges with an average degree of 12.63, however, the median degree is just 4. The largest connected component consists of 35,716 nodes and it is depicted in Fig. 15.

The degree distribution of the network is illustrated in Fig. 12. There are 897 isolated nodes in the graph (nodes with zero degrees), i.e. scholars who have a single-authored network science paper but have not co-authored any network science papers. The most typical number of co-authors are between 2 and 4 and the tail of the distribution decays much slower than the number of authors per paper does (c.f. Fig. 5) since here the degree reflects all the number of co-authors who do not necessarily author the same paper. The highest degree is 546 corresponding to Roberto Bellotti, a medical physicist, who is also an author of the paper with the highest number of collaborating authors [18] and another many-authored paper [9]. While our network is unweighted by definition, a possible weight could be assigned to the edges corresponding to the number of joint papers written by the two authors
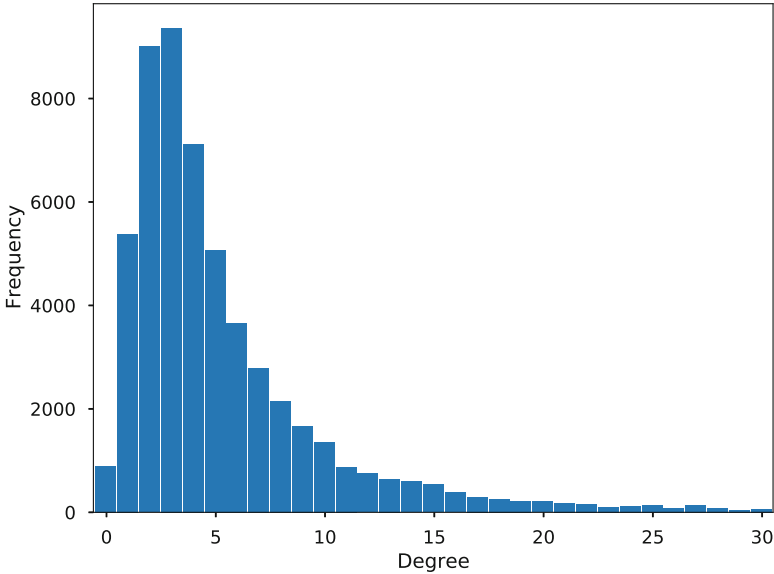
Fig. 12 The degree distribution of the network (truncated at 30)

Table 2 The most active links between authors

Authors		Number of joint papers
Shlomo Havlin	Eugene H. Stanley	52
Bing-Hong Wang	Tao Zhou	51
Jihong Guan	Shuigeng Zhou	50
Zhongzhi Zhang	Shuigeng Zhou	48
Jihong Guan	Zhongzhi Zhang	40
Zeng-Ru Di	Ying Fan	34
Sergey Dorogovtsev	José F.F. Mendes	32

at the endpoints of the edge. Table 2 shows the most 'active links', i.e. the edges with the highest weights in the edge-weighted version of the co-authorship network.

The network has a high assortativity coefficient of 0.53 that suggests that nodes tend to be connected to other nodes with similar degrees. The co-authorship network is highly clustered with a global clustering coefficient of 0.97 and an average local clustering coefficient of 0.8. The fact that the average shortest path length in the largest connected component is 6.6 also supports the small-world nature of co-authorship networks.

To identify the most central authors of the network science community as seen through the co-authorship network, we calculate centrality measures such as betweenness, harmonic and degree centralities of the nodes. The most central authors are shown in Table 3. We also compare the centrality measures of the authors with the citation count of their network science papers and with their *h*-

Table 3 The top 12 authors with the highest betweenness centrality. Their ranks with respect to each metric are shown in brackets

Name	Centralities			Number of citations	h-index
	Betweenness	Harmonic	Degree		
Jürgen Kurths	**0.025 (1)**	**0.169 (1)**	216 (1,017)	9,249 (30)	96 (2)
H. Eugene Stanley	0.024 (2)	0.168 (2)	220 (1,013)	10,479 (18)	57 (7)
Guanrong Chen	0.019 (3)	0.165 (4)	215 (1,018)	12,859 (15)	27 (30)
Albert-László Barabási	0.017 (4)	0.160 (12)	202 (1,023)	**73,937 (2)**	83 (3)
Yong He	0.014 (5)	0.163 (6)	**242 (1,012)**	9,104 (32)	61 (6)
Zhen Wang	0.014 (6)	0.163 (5)	155 (1,117)	3,306 (365)	39 (16)
Santo Fortunato	0.013 (7)	0.160 (13)	208 (1,021)	13,923 (12)	40 (14)
Shlomo Havlin	0.013 (8)	0.163 (7)	165 (1,042)	13,377 (13)	**110 (1)**
Tao Zhou	0.013 (9)	0.167 (3)	220 (1,013)	9,911 (20)	40 (14)
Edward T. Bullmore	0.012 (10)	0.151 (49)	210 (1,020)	17,915 (7)	50 (10)
Wei Wang[a]	0.012 (11)	0.161 (10)	145 (1,178)	467 (1511)	14 (188)
Stefano Boccaletti	0.112 (12)	0.162 (9)	130 (1,179)	9,609 (21)	22 (58)

[a]Sichuan University

indices (restricted only to their network science papers). Common characteristics of the most central authors that they are famous, well-established researchers, moreover, they are typically active in more research areas forming bridges between subdisciplines. The highest betweenness and harmonic centralities correspond to Jürgen Kurths, German physicist and mathematician whose research is mainly concerned with nonlinear physics and complex systems sciences. As we mentioned before, the highest degree corresponds to Roberto Bellotti, a medical physicist. Mark Newman English-American physicist has the highest number of citations on his network science papers (77,418), while Shlomo Havlin, Israeli physicist is ranked first with respect to h-index.

Figure 14 illustrates the relationship between centrality measures of network scientists and the scientometric indicators of their network science papers. On the left it shows the number of citations against the vertex betweenness centrality, colored by the harmonic centrality; on the right one can see the h-index against the vertex betweenness centrality, colored by the degree. We can conclude that there is a positive correlation between the authors' central role in the co-authorship network and their scientometric indicators. Figure 13 shows the Spearman's rank correlation heatmap of the aforementioned measures indicating positive correlations, with the highest positive correlation between citation count and h-index. Considering centrality measures against scientometric indicators, betweenness centrality and h-index has the highest correlation (Figs. 13 and 14).

Network scientists have become more connected as time has gone by, as it is illustrated in Fig. 16, since not only the size of the largest component has increased over the years but also the ratio of the size of the giant component to the size of the entire network, indicating the emergence of a diverse but not divided network
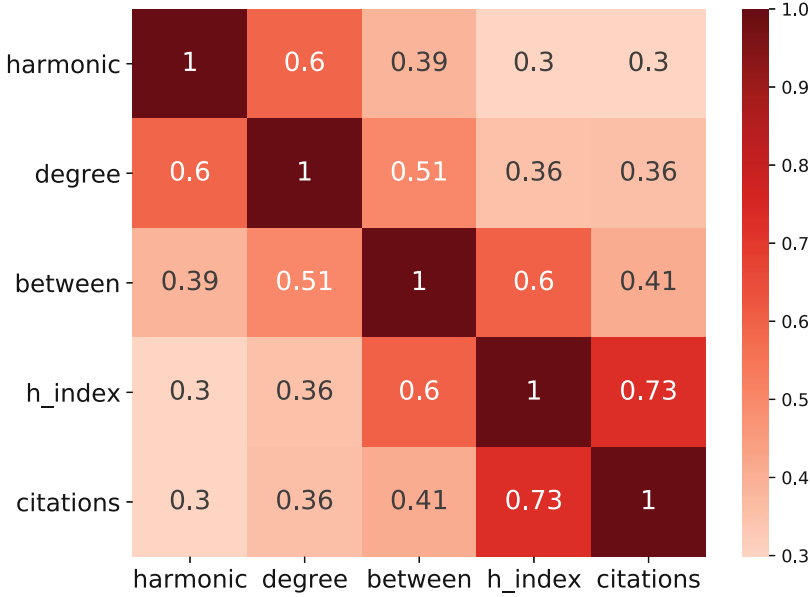
Fig. 13 Spearman correlation heatmap between various centrality measures and scientometric indicators
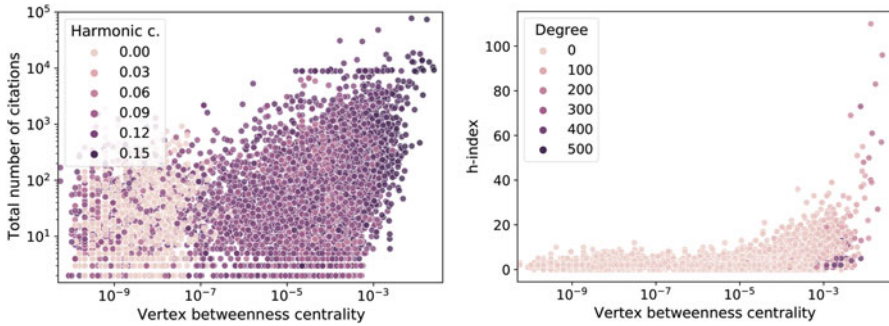

Fig. 14 Relationship between centrality measures of network scientists and the scientometric indicators of their network science papers

science community. The giant component consists of 35,716 nodes that is 63% of the entire network size and it is illustrated in Fig. 15.

Using Clauset-Newman-Moore greedy modularity maximization community detection algorithm [10], we identify the dense subgraphs of the network. To retrieve some important discipline and location-related characteristics of the largest communities, we assigned a research area and a country for each author as the majority of the research areas corresponding to their papers and the most frequent country of their affiliations respectively. The compositions of the ten largest