



Collaboration between authors in the field of social network analysis

Daria Maltseva¹ · Vladimir Batagelj^{1,2,3}

Received: 21 October 2021 / Accepted: 23 March 2022 / Published online: 19 April 2022
© Akadémiai Kiadó, Budapest, Hungary 2022

Abstract

This paper presents a study of authors writing articles in the field of SNA and groups them by means of bibliographic network analysis. The dataset consists of works from the Web of Science database obtained by searching for “social network*”, works highly cited in the field, works published in the flagship SNA journals, and written by the most prolific authors (70,000+ publications and 93,000+ authors), up to and including 2018. Using a two-mode network linking publications with authors, we constructed and analysed different types of collaboration networks among authors. We used the temporal quantities approach to trace the development of these networks through time. The results show that most articles are written by 2 or 3 authors. The number of single authored papers has dropped significantly since the 1980s—from 70% to about 10%. The analysis of three types of co-authorship networks allowed us to extract the groups of authors with the largest number of co-authored works and the highest collaborative input, and to calculate the indices of collaborativeness. We looked at the temporal properties of the most popular nodes. We faced the problem of “multiple personalities” of mostly Chinese and Korean authors, which could be overcome with the adoption of standardized author IDs by publishers and bibliographic databases.

Keywords Social network analysis · Bibliographic network · Temporal network · Co-authorship · Collaboration · Islands approach

✉ Daria Maltseva
dmaltseva@hse.ru

Vladimir Batagelj
vladimir.batagelj@fmf.uni-lj.si

¹ National Research University Higher School of Economics, 11 Pokrovsky Bulvar, Pokrovka Complex, Moscow, Russia 101000

² Institute of Mathematics, Physics and Mechanics, Jadranska 19, 1000 Ljubljana, Slovenia

³ University of Primorska, Andrej Marušič Institute, 6000 Koper, Slovenia

Introduction

The field of social network analysis (SNA) has rapidly developed over the past decades, which can be seen in the increasing number of scientific publications and the different disciplines where it is applied (Borgatti & Foster, 2003; Maltseva & Batagelj, 2019; Otte & Rousseau, 2002). According to the history of SNA development described by Freeman (2004), in the 1970s, the field could be represented by a set of scientific groups, unrelated to each other. However, by the beginning of the 1990s, SNA could already be defined as a “normal science”, in terms of Kuhn (Hummon & Carley, 1993), and its representatives formed an “invisible college” (Freeman, 2004), with a core groups of scholars having a shared paradigm and regular interaction. To a great extent, this development was connected with the field’s unification with the establishment of the *INSNA, International Network for Social Network Analysis* in 1977, the creation of the journals *Connections* in 1977 and *Social Networks* in 1978, conferences and regular meetings (including the annual conference *Sunbelt* first held in 1981), the appearance of computer programs standardizing network data analysis, educational programs, and the “bridging” work of some scholars. These led to the institutionalization of the field by the 1990s, when “*the representatives of each of these network “schools” [...] joined together and organized themselves into a single coherent field*” (Freeman, 2004, p. 135). Starting from the 2000s, the methods of SNA, which were mostly developed in social sciences, received considerable attention from natural and computer science researchers, and economists. This has been called “*the invasion of the physicists*” (Bonacich, 2004), who developed the network science (NS) discipline (Freeman, 2004, 2011). The development of the internet in the 1990s and online social networks in the 2000s increased the interest in networks, and extended the areas where SNA methods are applied. Some recent studies (Maltseva & Batagelj, 2019, 2020) show that currently SNA is represented not only by scholars from social sciences, physics, and computer science, but also many others, including neuroscience, medicine, and animal social network analysis in behavioral biology.

Over the last decades, various tools and metrics have been proposed and used extensively in bibliometric and sociological studies to characterize the development of scientific fields and their social and cognitive dynamics. In comparison to qualitative analysis, traditionally developed in the sociology of science, these metrics are based on the quantitative analysis of citations, authors, journals, keywords and other bibliometric units for the description of the modern science. We follow this tradition in this study, where we look at the current development of SNA through the analysis of collaboration structures of the scientific community of scholars involved in network studies. The observed scientific community is understood as authors publishing papers in the area of SNA. The analysis of collaboration in science is often based on an investigation of co-authorship and citation networks (however, it should be noted that besides these conventional forms there are other ways of scientific communication which may lead to collaboration in science—see, e.g. Katz & Martin (1997), Laudel (2002). Co-authorship networks are social networks, where a node is a social actor (an author), and citation networks are information networks, where a node is usually an artifact (a paper, a journal, or a keyword) (Yan & Ding, 2012). While the latter are important for depicting the whole structure of scientific knowledge, co-authorship networks detect research communities and identify collaboration patterns in academic disciplines.

The current study is aimed at the analysis of scholarly networks based on the relations of co-authorship. In these networks, nodes represent authors connected by an edge if they

have co-authored one or more publications; different types of normalization can be used (Batagelj & Cerinšek, 2013; Newman, 2001). Following the three approaches used to study scholarly networks (Yan & Ding, 2012), we pose the following research questions:

- On the macro-level, we are interested in the global structural (statistical) features of publishing activity and general trends of collaboration in the field of SNA.
- On the meso-level, we are interested in the most visible collaborative groups of authors that can be detected in the field under study.
- On the micro-level, we are interested in the most prolific authors involved in SNA and their individual behavior, including the changes through time.

This information can give us insights about the whole field's development, and see the range of disciplines where SNA evolves the fastest.

Various methods of bibliometric analysis have already been used to study the development of SNA. Scientific co-authorship structures in SNA, which are the focus of our analysis, were studied by Batagelj et al. (2014), Leydesdorff et al. (2008), Lietz (2009), Otte & Rousseau (2002). In the sense of the research of the field's development, attention was also given to citation and co-citation structures and bibliographic coupling for sets of works, journals, authors (Batagelj et al., 2014; Brandes & Pich, 2011; Chen, 2005; Hummon & Carley, 1993; Leydesdorff et al., 2008; Maltseva & Batagelj, 2019, 2021), and keyword co-occurrence networks (Leydesdorff et al., 2008; Maltseva & Batagelj, 2020). These topics were also discussed in studies of different subfields (Batagelj et al., 2014, 2020; Hummon et al., 1990; Kejžar et al., 2010) and subdisciplines within the field (Borgatti & Foster, 2003; Lazer et al., 2009; Otte & Rousseau, 2002; Varga & Nemeslaki, 2012). Some studies were oriented to the analysis of network science literature—the citation networks of publications for the knowledge domain of complex networks in general (Shibata et al., 2007, 2008, 2009), or “small world” literature (Garfield, 2004). More historiographically oriented works written by Freeman (2004, 2011) should also be mentioned, as well as the work of Hidalgo (2016) addressing the main differences between the streams advanced by social and natural scientists—“*two groups of academics that often fail to see eye to eye*”. However, providing important information about SNA, these studies are based on specific datasets, cover only selected subtopics in time periods which are no longer up-to-date, and provide only partial information which is difficult to compare.

We believe that the whole picture of the current development of SNA can be drawn only through the analysis of a dataset collected by a comprehensive search, which would include data from different areas of SNA. As a source of information in our study, we use the articles from the *Web of Science (WoS) database (Core Collection)* found with the query “social network*”, as well as other highly-cited works in the SNA field, and those published in the flagship SNA journals indexed in WoS, and written by the most prolific authors. Time coverage is up to and including 2018. The approach to bibliometric analysis used is based on the methodology proposed in previous studies of different scientific fields and topics (Kejžar et al., 2010; Batagelj et al., 2014, 2017, 2020). In the paper, this approach is extended by the analysis of the corresponding temporal networks (Batagelj & Praprotnik, 2016; Batagelj & Maltseva, 2020), which was previously applied to large bibliographic networks only partially. Together with a better understanding of the field under study, this article is intended to make a methodological contribution, presenting an example of the approach to temporal networks analysis developed before. The proposed approach to the analysis of collaboration structures can be also used in future studies aimed at the analysis of other scientific fields.

The rest of the paper is organized as follows. “[Previous studies on authors in SNA](#)” section presents previous studies of co-authorship among SNA authors. “[Data](#)” section describes the dataset and the construction of the derived networks from the original two-mode networks connecting works and authors/keywords. “[Results](#)” section provides the main results of the study. “[Distributions of works and authors in authorship network](#)” section presents the statistical properties of the authorship network, and a list of the most prolific authors with the largest number of works. “[Collaboration between authors](#)” section is based on the analysis of three collaboration networks, constructed using different types of normalization for the link weights. It shows the metrics of the total contribution to the field and collaborativeness indices for the most prolific authors, identifies the authors with the largest number of co-authored works, shows the groups of collaborating authors, and the keywords and journals associated with some of these groups. We use temporal versions of some of these networks to gain an insight into the dynamics of these relations, looking at the metrics for some selected well-known authors more precisely. We finish the article with general conclusions and a discussion of the results.

Previous studies on authors in SNA

Most of the works studying the structures of SNA authors are based on the analysis of co-authorship. Based on the data from the *Sociological Abstracts* database (1601 articles published 1963–2000), Otte & Rousseau (2002) constructed a co-authorship network of 113 authors occurring 3 times or more. The largest connected component of 57 authors, regarded as the most important scientists in the field, formed 16 cliques of 3 or more nodes. Several groups were identified in the network formed from such authors as Killworth and McCarty, Skvoretz, and Snijders. Doreian was the most central author in the sense of closeness and betweenness, and Wellman was the most prolific. Burt was an exception as he was not part of the main component: although he has published a large number of works most of them were written as single authored papers.

Leydesdorff et al. (2008) presented a temporal analysis of co-authorship networks, constructed from the publications in *Social Networks* in 1988–2007. Among the 445 authors published during this period, 364 co-authored with another, but only 146 of these co-authors were part of the largest component in any of the years under study, and were included in the analysis. The central part of the network consisted of the many well-known scholars in SNA methodology, who have contributed on diverse fronts for a long time, such as Freeman, Breiger, Everett, Borgatti, Pattison, Krackhardt, and D. White. The branches of the network represented more specialized developments in SNA. Some of these developments were more recent, such as the new statistical methods for SNA (ERGM—Handcock, Robins; the statistical estimation of longitudinal network data—Snijders; eigenvector centrality—Bonacich) and generalized blockmodeling (Batagelj, Ferligoj, Doreian), while other developments were older, such as the issues of network data collection (Bernard, Killworth, and McCarty) and the analysis of random graphs (Wasserman, Skvoretz, and Fararo). This last group was connected to the central part by other SNA pioneers, Faust and Carley. Wellman and colleagues formed a separate group. The analysis of temporal data showed that several authors, who were already in the center of the network (such as Freeman and D. White, or Borgatti and Carley from the second generation) moved around. Leydesdorff et al. regarded them as generalists combining work on network analytic methodologies in general with a specialization in substantive applications.

Batagelj et al. (2014) studied the collaboration networks among all the authors having papers on social networks (SN5 dataset, *WoS*, descriptions of articles before 2007, by Batagelj (2005)). The approach allows the construction of different collaboration networks and looks at the authors' personal characteristics—the largest number of collaborators and self-contribution scores to SNA. Batagelj et al. identified several strongly collaborating groups of authors: Borgatti and Everett (applying graph theoretic ideas to the general analyses of networks), Bernard, Killworth, McCarty, Johnsen, Shelley (working on informant accuracy and scale-up methods for estimating the size of hidden populations), Batagelj, Ferligoj, Doreian (blockmodeling), Rothenberg, Muth, Poterat, Woodhouse (networks for medical issues including the sexual transmission of disease), Magliano, Maj, Matangon, Fiorillo (networks for psychiatric interventions). The approach of Batagelj et al. clearly showed the existence of two large components: one represented by mainstream social network researchers (with such intensively collaborative pairs as Borgatti and Everett, Killworth and Bernard, Bonachich and Bienenstock, Ferligoj and Batagelj, Pattison and Robins), and the second consisting of physicists (with such pairs as Newman and Park, Barabasi and Albert, and Masuda and Konno).

In another study, Batagelj et al. (2020) used the dataset of publications in the area of graph and network clustering and blockmodeling (*WoS*, descriptions of articles before 2017) for the analysis of collaboration structures. Extracting the groups of the most connected authors (P_5 -cores) they found that social science researchers were the minority. The largest group was formed by the representatives of physical sciences—Newman, Barabasi, Peixoto, Fortunato, Lancichietti—working on the topic of community detection. Borgatti and Everett were connected to this group due to their work with Boyd. Other clusters were formed by Wasserman, Pattison, and Breiger (role systems, the foundations of ERGM), and Doreian, Batagelj, Ferligoj, Mrvar, Brusco, Steinley (algorithms for blockmodeling). Interestingly, the physicists Newman, Peixoto, and Turcotte did not appear on the list of the most collaborative authors, due to their publishing single author papers and publishing with many *different* co-authors.

In contrast to the mentioned studies, observing the field from the social network perspective, Lietz (2009) analyzed scientific co-authorship structures of the authors belonging to the research domain of the “new science of networks” (NSoN) based on literature on small-world, scale-free, and complex networks. The author showed that the field's social structure was becoming more dense continuously, forming a giant component, in which a large part of the entire network's authors collaborate at least indirectly. This allows to conclude that the development of the field satisfies the definition of emerging science, establishing a paradigm.

These findings show that the analysis of collaboration usually identifies a number of groups of scholars representing the field. The distinction of these groups is due to their affiliation to separate disciplines; the separation between the representatives of social sciences and network sciences, also observed in the previous studies of SNA (Bonacich, 2004; Freeman, 2011; Batagelj et al., 2014, 2020; Maltseva & Batagelj, 2019), is the most obvious. Within social sciences, researchers are also separated due to the subfields and topics (methodological and substantive) they are working on. Some of the groups of authors are represented by prominent researchers in the SNA field, who can be seen as the core group of the discipline. The results also highlight the importance of the dataset for the results: whether the analyzed articles are from social sciences (e.g., Otte & Rousseau, 2002) or they represent a topic where different disciplines are initially involved (e.g., Batagelj et al., 2020). We expect that through the analysis of a large and complete dataset, the representatives of other disciplines, not only social sciences and network science, can be identified.

Data

Data collection

The procedures of data collection, basic networks construction and cleaning were presented in detail before (Maltseva & Batagelj, 2019). Here we reproduce some essential information from this paper. The dataset consists of articles from the WoS database *WoS Core Collection*, Clarivate Analytics's multidisciplinary database of bibliographic information containing over 21,100 peer-reviewed, high quality scientific journals published worldwide in over 250 areas of science, including social sciences and humanities (Clarivate Analytics, 2020). Previous comparisons of different databases of bibliometric data, such as *Scopus*, *Google Scholar*, and special citation resources and scientific social media, such as *SciFinder*, *Mendeley*, have shown that they vary significantly according to their coverage of certain scientific disciplines, and have their pros and cons (Hilbert et al., 2015; Harzing & Alakangas, 2016; Martín-Martín et al., 2018). The *WoS* contains mainly publications from the journals with a certain impact factor (however, recently the conference proceedings and book data also became available) and provides coverage back to 1900 with bibliometric descriptions including references. Its higher consistency and accuracy of data and cover-to-cover indexing of the journals made the choice of the *WoS* most appropriate for the current study.

The initial dataset was formed from the publications matching the query “social network*”, and thus some works related to the broader field of network analysis could have been overlooked. The search query for “network analysis” would be too broad, including the works on computer networks, optimization problems for networks, etc. That is why we extended the results of the original query with a saturation search of papers which were intensively cited, and included the works from the flagship SNA journals indexed in WoS, and those published by the most prominent authors (around 100 scholars). The obtained dataset covers not only the works of social scientists, but also influential papers published by physicists, biologists, information and computer scientists, etc. The dataset covers the works published up to and partially including 2018.

Using **WoS2Pajek 1.5** (Batagelj, 2017), we transformed our data into a collection of linked networks. We constructed a one-mode citation network **Cite** on works (from the field CR of the WoS article description) and 3 two-mode networks: the authorship network **WA**, the journal network **WJ**, and the keyword network **WK**. In these two-mode networks, the first mode consists of works, and the second—authors (from the field AU), journals (from the field CR or J9), and keywords (from the fields ID, DE or TI). The works which have full WoS descriptions are called *hits*, and those which are cited only and are listed in CR field are called *terminal* works. For them, only partial information is provided: the *first* author only, journal, publication year, journal volume and the first page number. For some hundreds of the frequently cited terminal works we additionally included their complete descriptions without references.

The work's description (record) from the WoS is semi-structured. In many cases, different fields include a free text in which the entities (authors, journals, keywords, etc.) have to be identified. In the process of entity resolution we have to balance between synonymity (different terms denoting the same entity) and ambiguity (a single term denoting different entities). Our networks are dependent on the solutions provided in the program **WoS2Pajek**, and very time consuming manual improvements (checking some hundreds of the most frequent entities of each type, identifying equivalent journal

names in the data cleaning phase). Such improvements is not an option for large datasets; we have to do them automatically and assume that the errors can be considered as a kind of noise. We still have to check the obtained results carefully and in case that some error pops up we have to appropriately correct the data and rerun the analyses.

For work names, we used the *short names* of the following format: `LastNm[:8] + '_' + FirstNm[0] + '(' + PY+ ')' + VL + ':' + BP` (author's last name and first initial, year of publication, volume, beginning page). For example, `GRANOVET_M(1985)91:481`. For last names with prefixes the spaces are deleted, and unusual names start with characters * or \$. The names of the authors are encoded by the first 8 characters of their surnames and first initial of first names, such as `GRANOVET_M`. With this approach, some problems of author name recognition can occur. One problem is related to errors in data, when the names of same authors can be presented in a different way (e.g., `GRANOVET_M` and `GRANOVET_`). We identified such cases for all works with large indegree frequencies in the network **WA**, making an equivalence partition of nodes and shrinking the set of works accordingly in all the obtained networks (Batagelj et al., 2014, pp. 395–399). Another problem is author disambiguation, when different authors have the same names, well-known in the literature as the problem of “multiple personalities”(Harzing, 2015). It is especially relevant for authors with Chinese and Korean names due to the “three Zhang, four Li” https://en.wikipedia.org/wiki/List_of_common_Chinese_surnames effect, but can occur also with authors with common surnames (e.g., Smith, Rodriguez, Johnson). For such authors, the solution of **WoS2Pajek** does not perform well: different authors, having the same surname and first initial of their first name, merge during the creation of the network **WA**. This problem can be overcome if we use a special ID (such as ORCID) for each scientist. Unfortunately, this information is not provided in all WoS descriptions. We have to accept this as a limitation of the study.

We produced *basic* networks **CiteN**, **WAn**, **WJn**, and **WKn**, with the sets of the sizes shown in Table 1. Multiple links and loops were removed. As terminal works contain information on the first author only, it is not correct to use basic networks for the analysis of the connections between authors. For further analysis, we constructed *reduced* networks on hits **CiteR**, **WAr**, **WJr**, and **WKr**, the sizes of which are also shown in Table 1.

Table 1 Sizes of original and reduced (marked in bold) networks

	# Mode 1	# Mode 2
CiteN	1,297,133	
CiteR	70,792	
WAn	1,297,133	395,971
WAr	70,792	93,011
WKn	1,297,133	32,409
WKr	70,792	32,409
WJn	1,297,133	69,146
WJr	70,792	8943

Derived networks

In the analysis of networks we will use two powerful tools—the network normalization (fractional approach, Gauffriau et al., 2007) and the network multiplication (for details see Batagelj & Cerinšek, 2013; Batagelj et al., 2014, 2020b)—to construct *derived* networks.

Multiplying the two-mode network \mathbf{WA} by itself, we obtained the co-authorship or collaboration network $\mathbf{AA} = \mathbf{WA}^T * \mathbf{WA}$, where the weight of the edge between two authors $\mathbf{AA}[a, b]$ shows the number or share of works to which a and b both contributed. Normalizing the weights of the links (“[The normalization of derived networks](#)” section), different types of derived networks of collaboration (“[Different types of collaboration networks](#)” section) can be produced.

Multiplying the two-mode networks \mathbf{WA} and \mathbf{WK} , we can also construct the network of authors and keywords $\mathbf{AK} = \mathbf{WA}^T * \mathbf{WK}$, where $\mathbf{AK}[a, k]$ counts in how many works author a used the keyword k , and the authors and journals network $\mathbf{AJ} = \mathbf{WA}^T * \mathbf{WJ}$, where $\mathbf{AJ}[a, j]$ counts how many works of author a were published in journal j . Detailed information on these networks construction is provided in “[Keywords and journals for selected groups of authors](#)” section.

The normalization of derived networks

Let us consider the authorship two-mode network \mathbf{WA} . Its work outdegree is equal to the number of authors of a work, and the author indegree is equal to the number of works to which an author contributed. Normalization creates the network $n(\mathbf{WA})$, in which, assuming that each author contributed to the work equally, the weight of each arc is divided by the outdegree of initial node of this arc. The total contribution of each work w is equal to 1.

$$n(\mathbf{WA})[w, a] = \frac{\mathbf{WA}[w, a]}{\max(1, \text{outdeg}_{\mathbf{WA}}(w))} \quad (1)$$

The max in Eq. 1 takes care of works without authors ($\text{outdeg}_{\mathbf{WA}}(w) = 0$).

A similar normalization of collaboration links, but with $\text{outdeg}_{\mathbf{WA}}(w) - 1$ instead of $\text{outdeg}_{\mathbf{WA}}(w)$ was proposed by Newman (2001), who interpreted the weight of the link between two authors in a “strict” way, as the proportion of time spent for the collaboration only with **other** co-authors (that is why “self-collaboration” in single authored papers is not taken into account).

$$n'(\mathbf{WA})[w, a] = \frac{\mathbf{WA}[w, a]}{\max(1, \text{outdeg}_{\mathbf{WA}}(w) - 1)} \quad (2)$$

In the next subsection, we show how these types of normalizations are used to construct different types of collaboration networks.

Different types of collaboration networks

There are different ways to create one-mode networks of co-authorship or collaboration between authors \mathbf{AA} out of the two-mode authorship networks (Batagelj & Cerinšek,

2013; Batagelj et al., 2014). Using **WA**r network, including 70,792 works and 93,011 authors, and different types of normalization, we created three collaboration networks—**Co**, **Cn**, and **Ct'**.

The easiest way to obtain the co-authorship network is to construct the *first collaboration network Co* (Batagelj & Cerinšek, 2013) by the multiplication of a transposed network \mathbf{WA}^T with the original **WA** network:

$$\mathbf{Co} = \mathbf{WA}^T * \mathbf{WA} \tag{3}$$

In the derived **Co** network, the weight of the edge between the nodes *a* and *b* is equal to the total number of works authors *a* and *b* wrote together. The weights of loops in **Co** are equal to the total number of works that each author has (which are also equal to the author indegree value in the **WA** network).

Second collaboration network Cn (Batagelj & Cerinšek, 2013) partially uses the *fractional approach*, where the contribution of authors to their own works and works with co-authors is considered. The network is constructed by the multiplication of a transposed original network \mathbf{WA}^T with the normalized network $n(\mathbf{WA})$:

$$\mathbf{Cn} = \mathbf{WA}^T * n(\mathbf{WA}) \tag{4}$$

In the derived network **Cn**, the weight of the edge (*a* : *b*) between the nodes (authors) is equal to the contribution of author *a* to works that he or she wrote together with author *b* (which can be asymmetric). Then the *author's total fractional contribution* to all his/her works is the value of the diagonal (loops) of the **Cn** network for a selected author. Based on this, Batagelj & Cerinšek (2013) proposed the *self-sufficiency index* S_a as the proportion of author's contribution to all his/her works $\mathbf{Cn}[a, a]$ and his/her total number of works (which is the indegree of the author *a* in the **WA** network), and the *collaborativeness index* K_a , which is complementary to it (and is equal to 1 minus self-sufficiency):

$$S_a = \frac{\mathbf{Cn}[a, a]}{\text{indeg}_{\mathbf{WA}}(a)} \tag{5}$$

$$K_a = 1 - S_a \tag{6}$$

Multiplying these two normalized networks, the *third collaboration network Ct* was constructed by the multiplication of a transposed regularly normalized $n(\mathbf{WA})^T$ network with the normalized $n(\mathbf{WA})$ network (used for producing the network **Cn**):

$$\mathbf{Ct} = n(\mathbf{WA})^T * n(\mathbf{WA}) \tag{7}$$

In network **Ct**, the weights of the edges between the nodes (authors) are symmetric; the weight $\mathbf{Ct}[a, b]$ is equal to the total fractional contribution of collaborating authors *a* and *b* to works they wrote together. The total contribution of a complete subgraph corresponding to a work is 1. The total contribution of an author is equal to the sum of the weights of all the works he or she authored (including single-authored papers, the contribution to which is stored in loops)—*author's contribution to the field*.

As an alternative, using Newman's type of normalization (Newman, 2001), the collaboration network **Ct'** was constructed by the multiplication of a transposed regularly normalized $n(\mathbf{WA})^T$ network with the Newman normalized $n'(\mathbf{WA})$ network:

$$\mathbf{Ct}' = n(\mathbf{WA})^T * n'(\mathbf{WA}) \quad (8)$$

The obtained network \mathbf{Ct}' is undirected and without loops. Its diagonal is set to 0. The total contribution of the complete subgraph corresponding to each work with at least two authors is 1. The weight $\mathbf{Ct}'[a, b]$ of the edge ($a : b$) is equal to the total fractional contribution of the “strict collaboration” of authors a and b to works they wrote together.

Temporal collaboration networks

A temporal quantity describes changes of selected property through time and provides an insight into its dynamics. They are a basis for the longitudinal approach to the description and analysis of temporal networks. An alternative approach is a cross-sectional approach in which a network is split into/represented by time slices (snapshots) that are analyzed separately and afterward combined into the final result (Holme & Saramäki, 2019).

Applying the *temporal quantities* approach (Batagelj & Maltseva, 2020; Batagelj & Praprotnik, 2016) to the network \mathbf{WA}_r , we constructed the corresponding temporal networks, using Python libraries Nets (<https://github.com/bavla/Nets>) and TQ (<https://github.com/bavla/TQ>) (Batagelj, 2020a). They are of two types—the instantaneous network \mathbf{WA}_i and the cumulative network \mathbf{WA}_c .

Using the normalization and multiplication of temporal networks, temporal versions of collaboration networks can be constructed. For our purposes, out of networks \mathbf{WA}_c and \mathbf{WA}_i , we constructed the two versions of the second collaboration network \mathbf{CnTi} and \mathbf{CnTc} , in a similar way to the second collaboration network \mathbf{Cn} :

$$\mathbf{CnTi} = \mathbf{WA}_i^T * n(\mathbf{WA}_i) \quad (9)$$

where

$$n(\mathbf{WA}_i)[w, a] = \mathbf{WA}_i[w, a] / \text{outdeg}_{\mathbf{WA}}(w) \quad (10)$$

and

$$\mathbf{CnTc} = \mathbf{WA}_c^T * n(\mathbf{WA}_c) \quad (11)$$

where

$$n(\mathbf{WA}_c)[w, a] = \mathbf{WA}_c[w, a] / \text{outdeg}_{\mathbf{WA}}(w) \quad (12)$$

In these networks, the weight of the edge ($a : b$) between the nodes (authors) is equal to the temporal contribution (including the author’s self-contribution) of authors a and b to works.

Results

In this section, the main results of the study are provided. “Distributions of works and authors in authorship network” section introduces the statistical properties of the authorship network, and a list of the most prolific authors with the largest number of works. “Collaboration between authors” section, based on the analysis of three collaboration networks, shows the metrics of the total contribution to the field and collaborativeness indices for the most prolific authors; identifies the authors with the largest number of co-authored works;

and shows the groups of collaborating authors. The analysis of the keywords and journals associated with some of these groups describes authors in terms of the disciplines where they use SNA methodology. We use temporal versions of these networks to gain an insight into the dynamics of some of these relations, looking at the metrics for some selected well-known authors more precisely. The results can be ascribed to three levels to study scholarly networks—macro, meso, and micro.

Distributions of works and authors in authorship network

Based on the distributions of works and authors in the network **WAR**, in this subsection we observe the main statistical properties of this network, look how the number of authors per work is changing through time, and extract the lists of the most prolific authors. Using the temporal versions of the networks **WAR** and $n(\mathbf{WAR})$, we trace the temporal distributions of the productivity and activity for the 10 most prolific authors through time.

The outdegree distribution of the number of authors per work in the network **WAR** (Fig. 1) shows that one fifth (13,157 or 19%) of all works are written by a single author, while half of all works have two (18,635 or 26%) or three (16,661 or 24%) authors. In our clean dataset, there are only 44 works to which the authors were not assigned. For some works, the number of authors is very large. The extreme case is the work *Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking* published in *Nature Biotechnology* in 2016, with 126 authors. Almost all works with a large number of co-authors are from the field of natural sciences (medical, health, epidemiological, and behavioral studies), where the inclusion of a large list of authors is a frequent practice. However, the third rated article *Discussion on the paper by Handcock, Raftery and Tantrum* published in *Royal Statistical Society Journal Series A: Statistics in Society* was written by 48 social network scientists discussing the basic paper.

Combining the partition of the number of authors in works with the partition of the number of works written by year, we obtained its temporal distribution, describing how the number of authors is changing over time. The results presented on Fig. 2 show that since the 1980s, the number of single-authored papers dropped from 68% to 12%. The number of

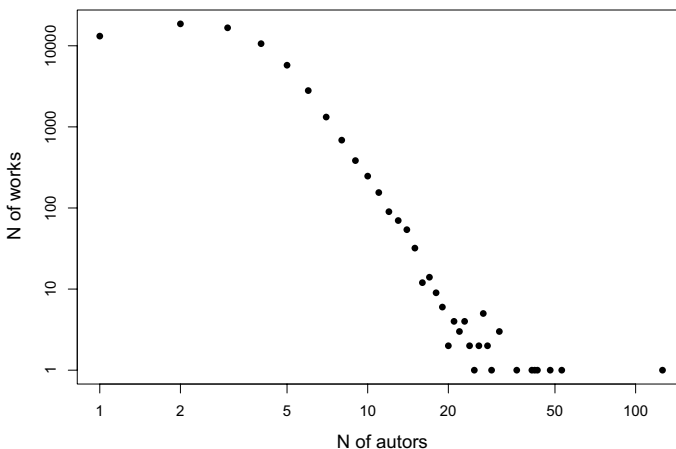
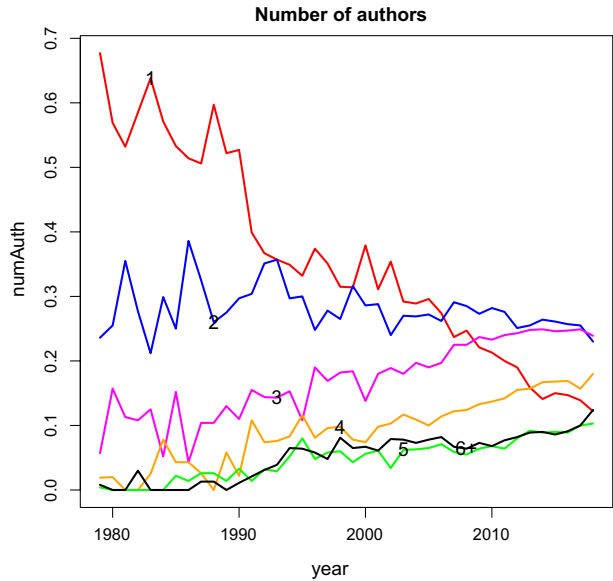


Fig. 1 **WAR**: outdegree distribution of number of authors per works

Fig. 2 Number of authors: temporal distribution of percentages



papers authored by two authors is relatively constant—around 25%. The numbers of papers authored by 3 or more authors are increasing (3: from 6% to 24%, 4: from 2% to 18%, 5: from 0% to 10%, 6+: from 0.8% to 12.4%). Besides the general trend of higher collaboration the reason could be also the expansion of SNA to other disciplines (physics, computer science, neuroscience, biology, chemistry, etc.) with different writing cultures.

In Fig. 3, the indegree distributions of the number of works per author in **WAR**—frequency and complementary cumulative—in double-logarithmic scale are shown. They fit the *power law* distribution $f = c \cdot n^{-\alpha}$, with fitted $\alpha = 2.53$ ($n_{\min} = 6$, using R package *powerLlaw*). This means that with the large number of articles published in the field, the authors producing many publications are less frequent (Lotka’s law). The complementary

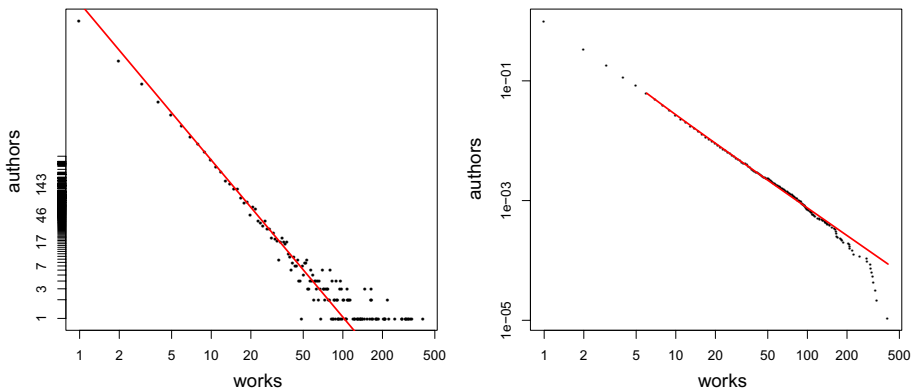


Fig. 3 **WAR**: indegree distribution of number of works per author—frequency (left) and complementary cumulative (right) in double-logarithmic scale

cumulative distribution confirms the scale-free nature of the most part of the indegree distribution, except for the most productive authors.

In the list of authors ranked in decreasing order of their **WAr** indegrees the top entries have Chinese or Korean surnames, e.g. (number of articles in parentheses): WANG_Y (410), WANG_X (339), ZHANG_Y (332), LIU_Y (321), CHEN_Y (317), ZHANG_J (310), LI_J (305), LI_Y (304), LI_X (287). The issue of the super-productivity of these (groups of) authors was discussed by Harzing (2015). This is also a partial reason that the tail of the complementary cumulative distribution in Fig. 3 is not following a straight line (fitting the power law). For example, Wang had almost 80 works published each year in the years 2015 and 2016.

Due to this problem, it was not productive to list the authors with the largest number of works. To extract the most prolific, but “real” ones, we had to ignore authors who had Chinese or Korean names. The first author that fitted this requirement was Latkin, ranked 45th, with 130 works, followed by Valente, ranked 72nd, with 97 works, and Dunbar, ranked 84th, with 91 works. Other most prolific authors with the largest number of works (top-60 authors with minimum 25 works) are presented in Table 2 (column “Total # of works”). However, with these names the authors disambiguation problem still occurs, as there are authors with such common surnames as Smith, Rodriguez, Johnson. We checked in secondary sources (Google, Google scholar) that the authors appearing in the top lists are single-person entities. In other cases, the table lists well-known authors from the SNA field. Many authors from the list represent *social* network analysis, including core SNA researchers, such as (in descending number of works) Latkin, Valente, Doreian, Carley, Burt, Borgatti, Snijders, Robins, Wellman, Pattison, Johnson, Martinez, Leydesdorff, Litwin, Bonachich, Everett, Contractor, White, Skvoretz, Moody, Marsden, M. Smith, Krachhardt, Thelwall, Friedkin, Wasserman, Brandes, Batagelj, Breiger, Scott, Freeman, Lazega, Faust. There are also authors representing the *network science* discipline emerged in the natural sciences—Newman, Barabási, González, J. Rodriguez, Pentland, Masuda, Watts. Other authors represent computer science (Kazienko, Faloutsos, Rahman, M. Rodriguez (Gomez-Rodriguez), Kleinberg), epidemiological and health studies (Christakis, Fowler, Schneider, Morris, Berkman), and behavioral biology (Croft, Farine).

For the detailed inspection, we selected the 10 most prolific authors from Table 2: Latkin, Valente, Dunbar, Newman, Christakis, Doreian, Carley, Burt, Borgatti, and Snijders. Using the networks **WAI** and **WAc**, we looked at their temporal distributions (instantaneous and cumulative) of the number of works (Fig. 4), where $\pi_{WAI}(a)$ is the *productivity* of author a —the frequency distribution of published papers of the author a per year.

$$\pi_{WAI}(a) = \sum_w WAI[w, a] \tag{13}$$

These distributions vary according to the periods, volumes, peaks and speed of publishing activity. Among these top-10 authors, Doreian and Burt have been present in the field from the start of their professional careers (and formation of SNA) in the 1970s; and Newman and Christakis were active only from the 2000s (which does not prevent them from having a similar number of works as everyone else). It is not possible to make any generalizations according to the selected set of authors, but these examples show that some authors can be constantly present in the field, from the start of their professional careers, while others are only temporarily active in the field of network analysis, which can be connected to their changing interests. In this sense, it is interesting to mention Newman, whose

Table 2 Network Cn: Authors' productivity and collaborativeness (the authors with the level of collaborativeness $\geq 50\%$ are marked in bold)

#	Author	Total # works	Total contribution	collaborative-ness, %	#	Author	Total # works	Total contribution	collaborativeness, %
1	LATKIN_C	130	32.99	75	31	EVERETT_M	44	22.58	49
2	VALENTE_T	97	34.96	64	32	MORRIS_M	43	17.22	60
3	DUNBAR_R	91	40.02	56	33	CONTRACT_N	43	14.15	67
4	NEWMAN_M	81	50.02	38	34	WHITE_H	42	27.28	35
5	CHRISTAK_N	74	22.89	69	35	SKVORETZ_J	42	20.07	52
6	DOREIAN_P	72	46.19	36	36	PENTLAND_A	41	14.12	66
7	CARLEY_K	72	28.11	61	37	MOODY_J	40	17.7	56
8	BURT_R	71	55.73	22	38	SMITH_A	40	14.2	65
9	BORGATTI_S	71	29.72	58	39	MARSDEN_P	39	30.17	23
10	SNIJDERS_T	67	29.63	56	40	BERKMAN_L	39	14.3	63
11	BARABASI_A	67	27.61	59	41	SMITH_M	39	13.29	66
12	FOWLER_J	65	20.14	69	42	KRACKHAR_D	38	18.24	52
13	KAZIENKO_P	64	21.97	66	43	JACKSON_M	38	17.78	53
14	ROBINS_G	64	19.67	69	44	THELWALL_M	37	18.41	50
15	WELLMAN_B	63	36.43	42	45	FRIEDKIN_N	36	28.17	22
16	FALOUTSO_C	60	17.86	70	46	SINGH_A	36	14.5	60
17	RAHMAN_M	59	19.18	67	47	WASSERMA_S	35	15.64	55
18	PATTISON_P	58	18.94	67	48	BRANDES_U	35	14.39	59
19	JOHNSON_J	54	21.19	61	49	GONZALEZ_A	35	14.13	60
20	MARTINEZ_M	53	21.9	59	50	KLEINBER_J	34	15.05	56
21	GONZALEZ_M	52	17.76	66	51	FARINE_D	34	14.04	59
22	RODRIGUE_J	52	15.9	69	52	BATAGEL_V	33	14.64	56
23	SCHNEIDE_J	52	13.89	73	53	BREIGER_R	31	19.73	36
24	LEYDESDO_L	51	33.28	35	54	WILLIAMS_A	31	14.5	53

Table 2 (continued)

#	Author	Total # works	Total contribution	collaborative-ness, %	#	Author	Total # works	Total contribution	collaborative-ness, %
25	LITWIN_H	50	32.42	35	55	SCOTT_J	28	17.54	37
26	RICE_E	48	13.09	73	56	MASUDA_N	28	14.26	49
27	BONACICH_P	46	34	26	57	FREEMAN_L	27	20.03	26
28	RODRIGUE_M	46	13.21	71	58	WATTS_D	27	13.67	49
29	NGUYEN_H	46	12.76	72	59	LAZEGA_E	26	14.17	46
30	CROFT_D	46	11.6	75	60	FAUST_K	25	13.5	46

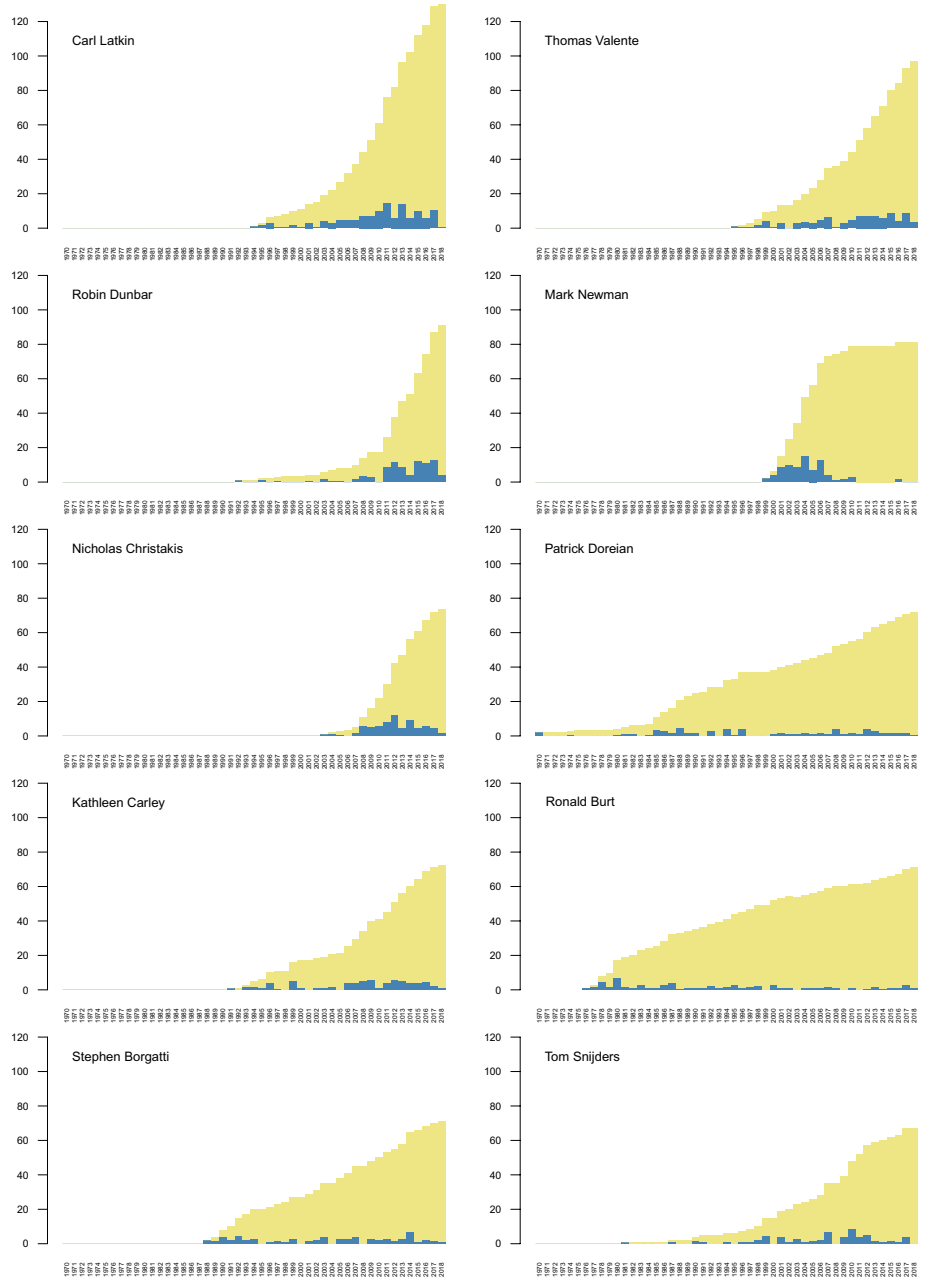


Fig. 4 Networks Wai and WAc: Productivity—number of works per years for selected authors, instantaneous (blue/dark) and cumulative (yellow/light)

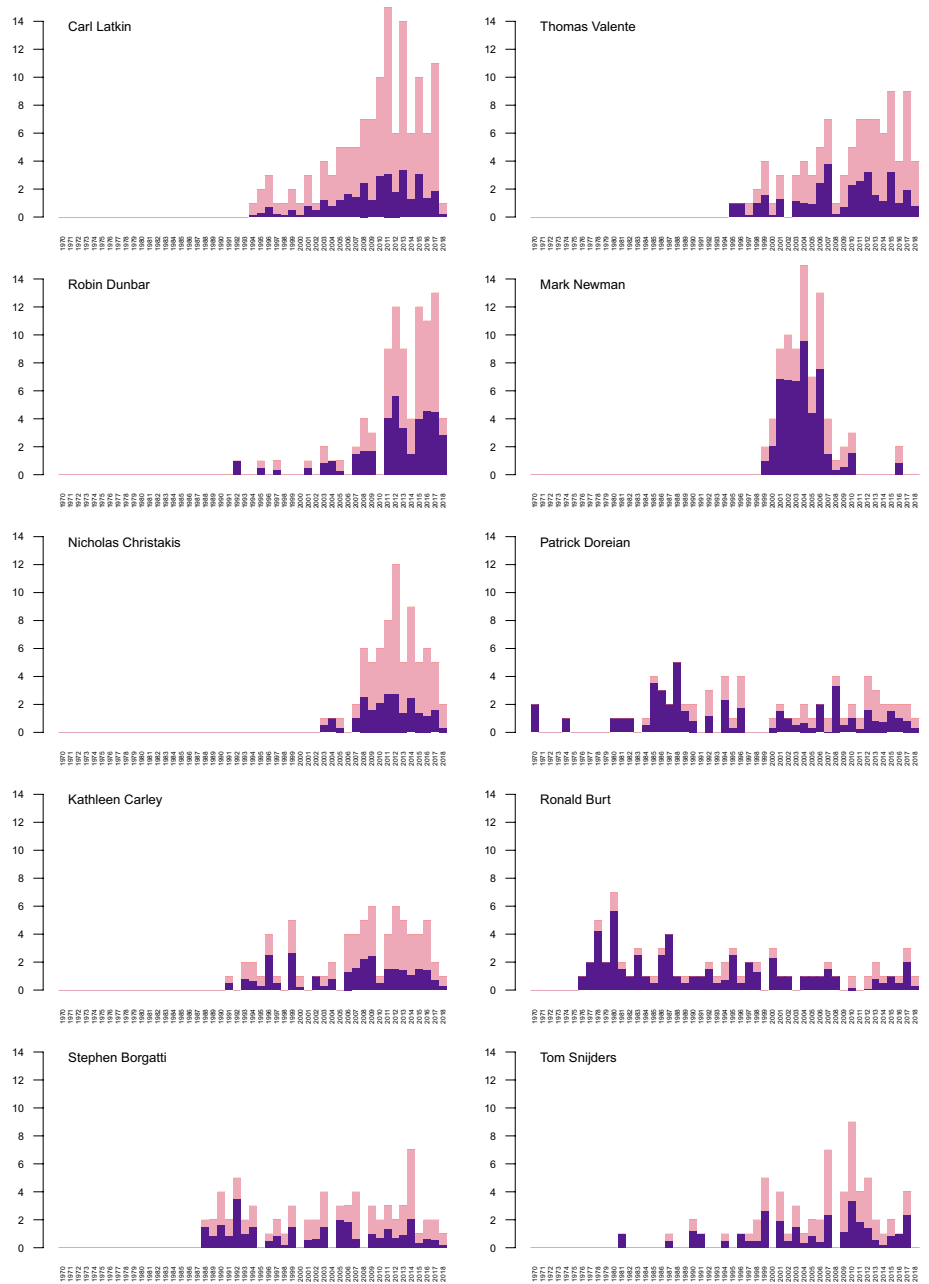


Fig. 5 Network WA: number of co-authored papers (pink/light) and fractional contribution (purple/dark) per years for selected authors

activity in SNA has been decreasing lately; maybe working on the second edition of Networks (Newman, 2018).

Figure 5 shows the temporal distributions of activity (number of co-authored papers based on the network **WAr**) $\pi_{\text{WAi}}(a)$ (pink/light) and fractional activity (fractional contribution to them, based on network $n'(\text{WA})$ constructed with the Newman normalization) $\pi_{n(\text{WAi})}(a)$ (purple/dark) for the top-10 authors. The distributions allow us to observe the differences between the number of works written by the authors and their contribution to them: while for some authors (e.g., Latkin, Valente, Dunbar, Christakis, Carley) these differences are quite significant, for others (e.g., Doreian, Burt, Newman) the measures of personal input are very close to the general ones (meaning that they work mostly by themselves).

Collaboration between authors

In this subsection, we present the results of the analysis of the three collaboration networks. First we show the individual characteristics of authors' contributions to the field and the indices of collaborativeness, including temporal ones for the 10 previously selected authors, based on the network **Cn**. Then we present the authors with the largest number of co-authored works, based on the network **Co**, and the groups of authors which can be identified in the network **Ct'**. Then we describe some of the identified authors' groups with the keywords and journals associated with them in order to understand their disciplinary identities.

Authors' contribution to the field and collaborativeness

It is interesting to compare the number of works that authors have published with their fractional contribution to these works and the level of collaborativeness, which can be obtained from the network **Cn**. We looked at these two measures for the top-60 authors presented in Table 2 (columns "Total contribution" and "collaborativeness"; the table is ordered by authors' total number of works). The authors with the indices of collaborativeness equal and larger than 50% are marked in bold.

For most of the authors with the largest number of works, their self-inputs (total fractional contributions) to their works are not very large. That is why the indices of collaborativeness for these authors are quite high—for Latkin, Christakis, Valente, Carley, Borgatti, Dunbar, who have more than 70 works, it varies from 56% to 75%. However, some of the authors with a large number of works have larger input into their own works and, accordingly, a low level of collaborativeness with others—22% for Burt, 36% for Doreian, and 38% for Newman. Similar results for some authors' orientation to single-authored papers were shown in Otte & Rousseau (2002), Batagelj et al. (2020). The authors with a large total number of works have two opposite patterns: some of them do not often collaborate, while others publish a large number of collaborative works. Among all the scientists listed in the table having large number of works published, the number of those who have the level of collaborativeness more than 50% is large.

Based on the temporal networks **CnTi** and **CnTc**, we calculated temporal collaborativeness instantaneous and cumulative indices for the top-10 authors from Table 2. Figure 6 shows cumulative (marked in pink/grey) measures, which are based on all previously published papers and that is why the measures are more stable, showing more general trends. The instantaneous (marked in yellow/light) measures are more variable from year to year,

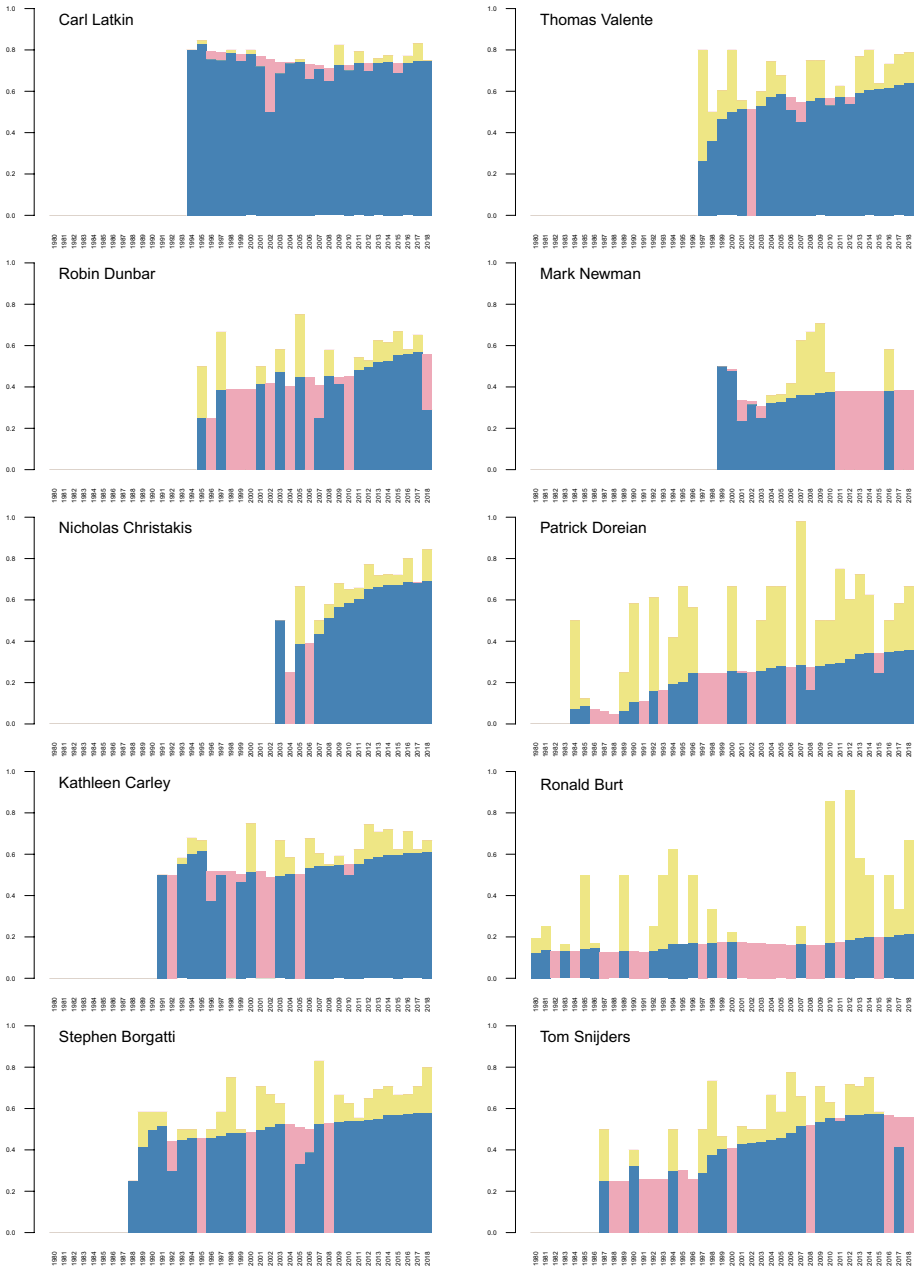


Fig. 6 Networks **CnTi** and **CnTc**: collaborativeness indices per years for selected authors—instantaneous (yellow/light), cumulative (pink/grey) and intersected (blue/dark) measures

as they show the collaborativeness in a particular year, or the short-term trends. The intersections of the two measures are marked in blue/dark. Again, not drawing any general conclusions, we can observe various possible patterns of collaborativeness. While for some

authors (e.g., Latkin) a high level of collaboration is always a case, for others (e.g., Valente, Christakis, Carley, Borgatti) this is more typical for the current stages of their careers. For others (e.g., Newman, Doreian, Burt) the high level of collaboration arises situationally, with gaps between the years. That is why their general and cumulative measures of collaborativeness are low (even though from time to time these authors are very active in collaboration).

Authors with the largest number of co-authored works

Due to the problem of “multiple personalities” it is not productive to look at the authors with the largest number of co-authors (which would be the indegree of the network *Co* from which the loops are removed). However, from this network we can get information about the authors having the largest number of works written jointly with someone else. We made a link cut at the level of at least 15 co-authored works and got a subnetwork of 114 nodes, which includes a component of 6 nodes, 2 components of 5 nodes, 4 components of 4 nodes, 8 components of 3 nodes, and 29 components of 2 nodes (Fig. 7). Half of the nodes (51%) belong to 2-node components. There are not so many authors who have 15 works written in collaboration, and even for them it is more common to work in pairs.

The structures represent groups of authors working in different disciplines and on various network-related topics. Some of these structures are complete subgraphs, where everyone is linked to everyone: the group of Khadilkar, Kantarcioglu, Thuraisingham, Khan, Abrol, Heaterly, working in the field of online social networks and social media (a minimum of 22 works written together); Kimura, Saito, Ohara, and Motoda, working in the field of artificial intelligence (minimum 23 works); Lax, Buccafurri, Nocera working in the field of online social networks (minimum 19 works), Vassilev, Rogers, Kennedy working in the field of medical studies (minimum 17 works); James, Croft, Krause working in the field of animal and human behavior (minimum 17 works); Potterat, Muth, Rothenberg working

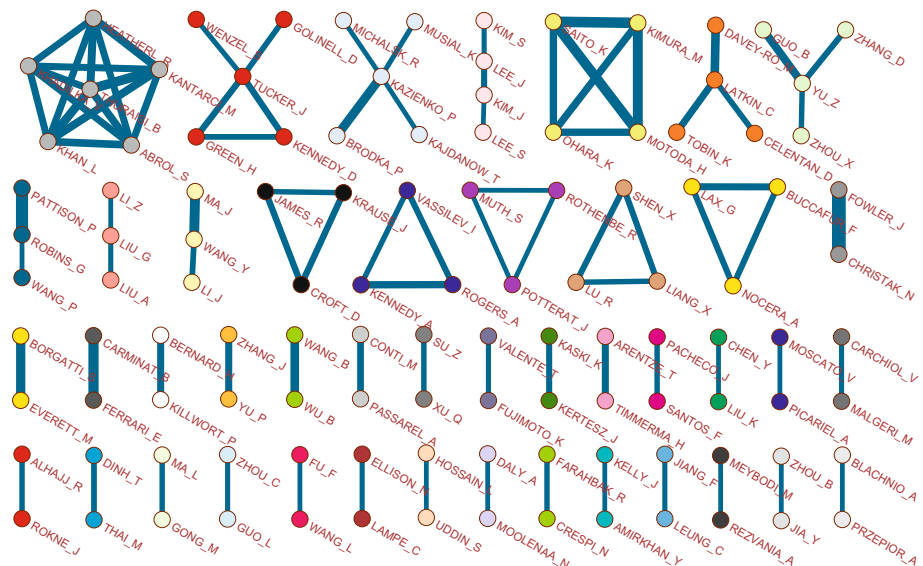


Fig. 7 Network *Co*: selected components

on risk and disease social networks (minimum 15 works). Several groups represent more star-like structures: a group around Latkin with minimum 18 works with Celentan, and a maximum 27 works with Davey-Rothwell, in the field of psychology and medicine; a group around Kazienko, having maximum 28 works with Bródka analyzing complex networks; and a group of Kennedy, Green, Tucker, Golinelli, Wenzel, working in the field of family, sex, and social support (minimum 16 works). Among the 29 pairs, there are authors with maximum values (the number of works written in brackets)—Fowler and Cristakis, working in health studies (43), Carminat and Ferrari—computer science (32), Borgatti and Everett (29), Killworth and Bernard (22), and Valente and Fujimoto (16)—SNA methodology, and Maybody and Rezvania (15)—complex networks methodology. Among the groups of 3 serially connected authors, the pair of Robins and Pattison appears with 38 works, connected to Wang with 15 works written together (ERGM model).

There are some special cases—groups of authors resulting from the effect of “multiple personalities” (Harzing, 2015). The largest link weights are for MA_J and WANG_Y, who have 31 works written together. According to the database, the author with the name MA_J has published works in computer science, social networks, energy, physics, and health. The given name of this author varies: *Jing, Jun, Jiemin*, etc. The given name of WANG_Y also varies: *Wang YZ, Wang YC, Wang YW*, etc. Other such pairs with a large number of papers in common are WANG_B and WU_B (27), GUO_B and YU_Z (25).

Groups of collaborating authors

We used the network Ct' to extract groups of authors collaborating with each other. Due to the problem of “multiple personalities”, we were unable to apply the usual way to identify the most collaborative groups of researchers— P_s —cores, as the Chinese and Korean groups of authors blurred the results. We used a *link islands approach* (simple and general islands) (Batagelj et al., 2014, pp. 54–57)—a very general and efficient approach to determine the “important” subnetworks in a given weighted network. For a given weight on links, a *link island* is a connected subnetwork having higher internal cohesion than links to its neighbors—it contains a spanning tree in which the smallest weight is larger than the largest weight linking the island to its neighbors. Usually we are searching for maximal islands of size (number of nodes) in a given interval $[k, K]$. An island is *simple* if it has a single “peak”. Simple islands contain authors working on the same topic.

Setting different lower and upper bounds for the size of islands, a different number of islands can be identified. Table 3 shows different results obtained for various bounds of simple islands: $[2-50]$, $[5-50]$, $[10-50]$, and $[20-100]$. The islands have a nested structure—for example, the bounds for 20–100 nodes creates only 3 islands consisting of 35, 23 and 21 nodes, which also appear for all other bounds. To preserve as much information as possible, we decided to use the boundary $[2, 50]$, which generated 14,222 islands. Besides the three mentioned largest islands, there are 70 islands having between 12 and 19 nodes.

Table 3 Network Ct' , simple islands: different ranges

#	Bounds	# islands	# nodes	% nodes
1	[2–50]	14,222	45,524	45
2	[5–50]	2192	14,215	15
3	[10–50]	173	2064	2.2
4	[20–100]	3	79	0.1

The largest part of the network (77%) consists of groups of relatively small sizes—5 (10%), 4 (15%), 3 (24%), and 2 (28%) nodes. The sizes and number of islands show that there are many groups of collaborating authors that can be extracted out of the network Ct' , and most of them consist of a relatively small number of authors.

There are different ways to identify “interesting” islands for the further investigation: to look at the islands (1) by size (consisting of many authors), (2) with the largest values of link weights (consisting of authors with strong ties with each other), (3) containing specific names (well-known authors in the field of SNA). Below, we used these three approaches to select the islands.

Largest islands Using the first approach, we extracted the 74 largest islands consisting of 12 to 35 nodes (1037 nodes, 2.2% of network). Part of these structures are not very interesting: they are star-like networks, which represent one author collaborating with many others, or (almost) complete clusters (cliques), where all authors collaborate with (mostly) everyone else. However, islands having non-trivial structures can be interesting. Figure 8 presents 34 selected islands (501 nodes). Among them, we identified the groups around some authors from Table 2—of physicists (Newman, Clauset, Girvan, Watts, Strogatz, Kossinets, Park) and social gerontologists (with Litwin in the middle). However, the identification of other authors needs some special search.

The problem of the approach based on the largest islands extraction is that the values in some largest islands can be very low, and in this sense the extracted structures can be not so important—the authors can be connected to each other and form a large group, but their collaboration might not be intensive. In our case, the link weights in the network Ct' varies between 0.0001 and 24.8333. For the largest islands obtained above, the ranges of the link weights are: island of 35 nodes—[0.018–2.00], 23 nodes—[0.035–2.202], 21 nodes—[0.005–1.00], and 19 nodes—[0.022–0.330]. It means that other approaches for island selection can be more productive.

Islands with the largest top link weights To get islands with strong links from the network Ct' , we made a link cut at the level of 7.5 and extracted a network of 32 nodes. Then we manually searched for the islands to which these 32 nodes belong, and extracted these islands (Fig. 9). Some of the authors with the largest link weights between each other had already appeared as the result of C_0 network link cut. These are the groups around Kimura, Saito, Ohara, and Motoda (artificial intelligence); Latkin and Davey-Rothwell (psychology and medicine); and larger groups of those who appeared as pairs before—of Borgatti, Everett, Boyd and Halgin (SNA methodology and UCINET), Fowler, Cristakis, and Shakya (health studies), Carminati, Ferrari, et al. (computer science). There are also groups connected to Barabási and Posfai, and Litwin and Stoeckel (who also appeared as the largest island above). Several groups are formed around Steinhausen and Metzke working in the field of medicine, physicists Grabowski and Kosiński working on artificial neural networks, and representatives of urban studies Arentze and Timmermans. There are also several groups of authors with Chinese and Korean names.

Islands for selected authors We decided to look further at the islands to which the authors from Table 2, who did not appear in largest islands and islands with largest link weights, belong. We manually checked the numbers for 47 authors and their simple islands. 5 authors (NGUYEN_H, WHITE_H, BREIGER_R, SCOTT_J, and FREEMAN_L) did not form an island in the network Ct' . The islands for the other 42 most prolific authors are presented in Fig. 10. The disciplinary attributes of these authors were already given above. Several groups have a relatively large number of authors—9 and more (around Williams, Masuda, A. Gonzalez, Dunbar, Smith); there are relatively many groups consisting of two and three authors. Again, some of these subnetworks are star-like structures (as the

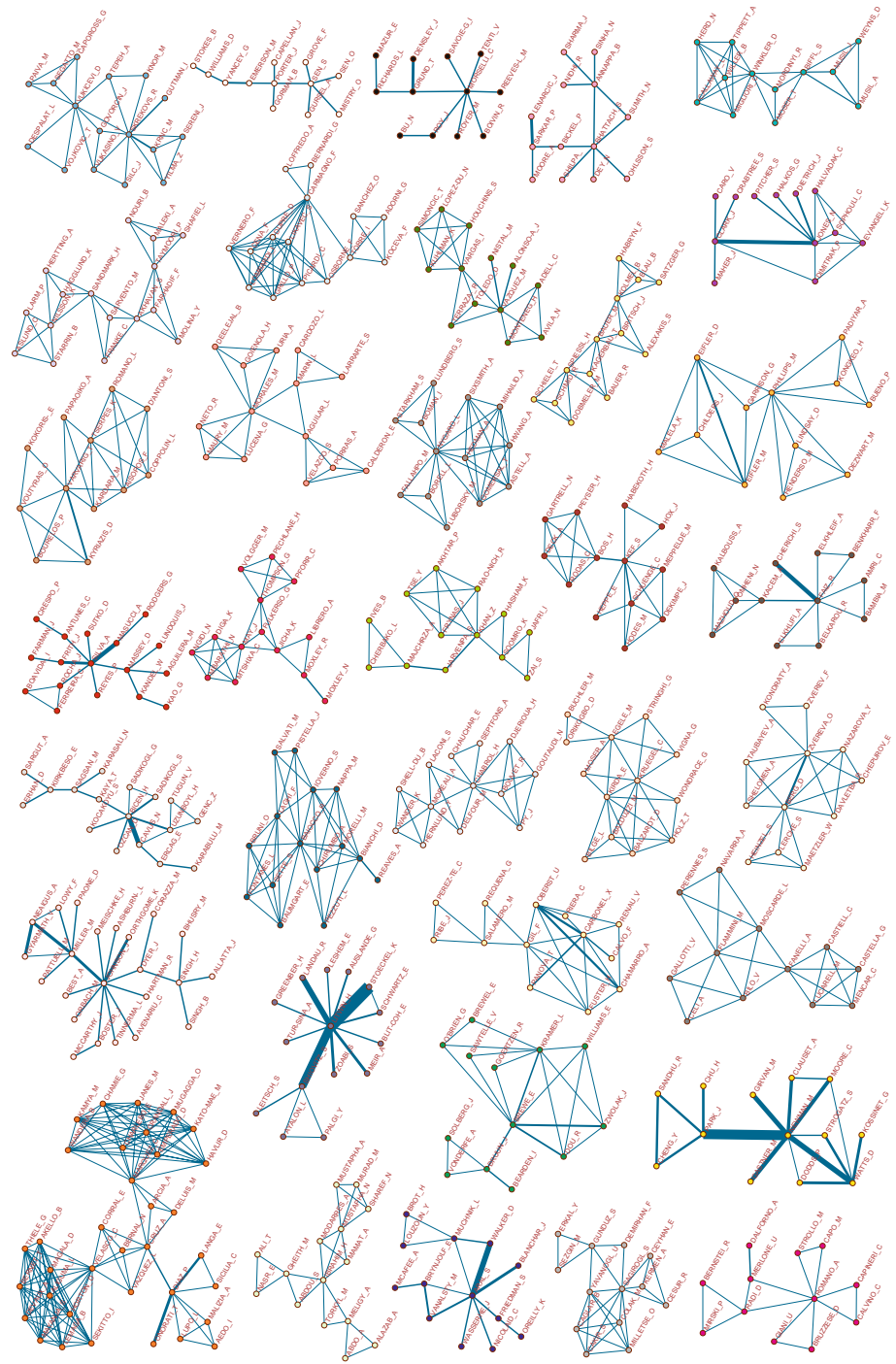


Fig. 8 Network Ct': selected islands

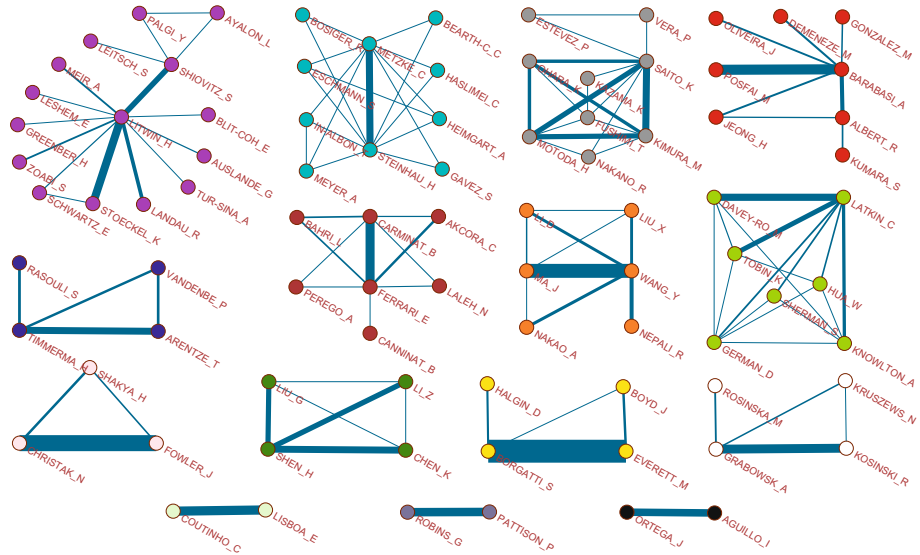


Fig. 9 Network Ct': simple islands for authors with the largest link weights

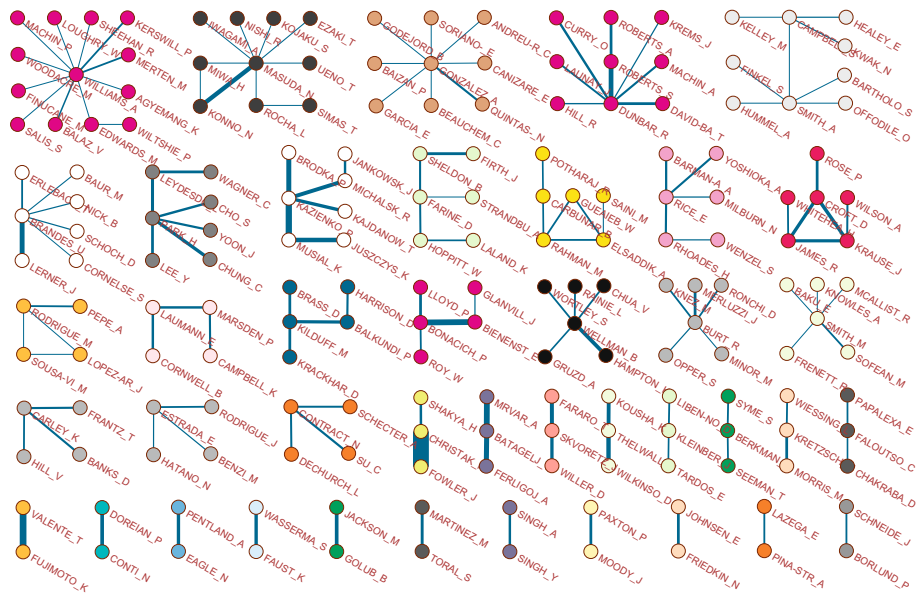


Fig. 10 Network Ct': simple islands for selected authors

ones around Williams, Masuda, Dunbar, Brandes, Wellamn, Burt, Smith, Carley, Estrada, Contractor).

Simple islands forming larger islands It is interesting to see how the groups of authors obtained by the simple islands approach from the network Ct', originally not highly linked

Table 4 Network Ct' , general islands: different ranges

#	Bounds	# islands	# nodes	% nodes
1	[2–100]	13,182	48,029	52
2	[2–50]	13,200	47,991	51.5
3	[10–50]	514	7374	8
4	[20–100]	70	1971	2.2

to each other, can form larger groups—authors working on similar topics. To find such structures, we used the *general islands* approach, which again resulted in a different numbers of islands (Table 4). For each boundary, this approach creates clusters with a larger number of nodes than the simple islands approach. For further investigation, we have chosen the boundary [2–100], which resulted in 13,182 islands having 48,029 nodes (52%). As with the previous approach, the largest part of the network (67%) is formed of islands of relatively small sizes—2 (24%), 3 (21%), 4 (13%) and 5 (9%) nodes. The islands of 20 nodes or more are formed only by 4% of all nodes. The largest two islands contain 96 and 80 nodes. The first island is formed of the authors with Chinese and Korean names; the second island presents many authors connected to each other—however, these are not the “core” or “classic” participants in SNA.

Again, as the inspection of all the islands is time-consuming and all the results cannot be presented in the article, we manually searched for the general islands formed out of several simple islands, for the 60 authors from Table 2. For these authors, there are 566 nodes in the subnetwork of general islands, in comparison to 189 nodes in the case of the simple islands. Figure 11 presents some of these islands. Most of the islands increased in size, attaching new nodes, and some have been merged from several simple islands. The largest change happened to the simple islands represented by Snijders, Skvoretz, Doreian, and Batagelj, and the pairs of Wasserman and Faust, and Robins and Patisson, who with Breiger (who did not form a simple island) form a joint SNA island of 42 nodes. Several other merged islands are also from social sciences; they are formed by the groups of authors connected to Carley and Krachardt, Leydesdorff and Thelwall, and Marsden and Schneider. Interestingly, other authors from SNA did not merge to these general islands and formed their own groups—of a larger size (e.g., represented by Dunbar, Valente, Moody, Borgatti and Everett) or the same size (e.g., represented by Brandes). Authors from other disciplines follow the same pattern and either enlarged their islands (e.g., represented by Masuda), or remained the same (e.g., represented by Newman, Kazienko, Bonacich). Thus, many “classic” representatives of SNA tend to be connected to each other, while other authors working on separate topics within the same discipline, or other disciplines, form separate groups of different sizes.

Keywords and journals for selected groups of authors

To uncover the research topic and disciplinary identities of a selected group of authors, from the normalized reduced networks WAr , WKr , and WJr , we constructed networks of authors and keywords $AK'' = n(WA)^T * n(WK)$, and authors and journals $AJ'' = n(WA)^T * n(WJ)$. Both networks are normalized. In the network AK'' , the weight $AK''[a, k]$ is equal to the fractional use (publishing in) of author a of keyword k . In the network AJ'' , the weight $AJ''[a, j]$ is equal to the fractional use of author a of journal j . For a given keyword k and journal j , it can be extended to a group of authors C .

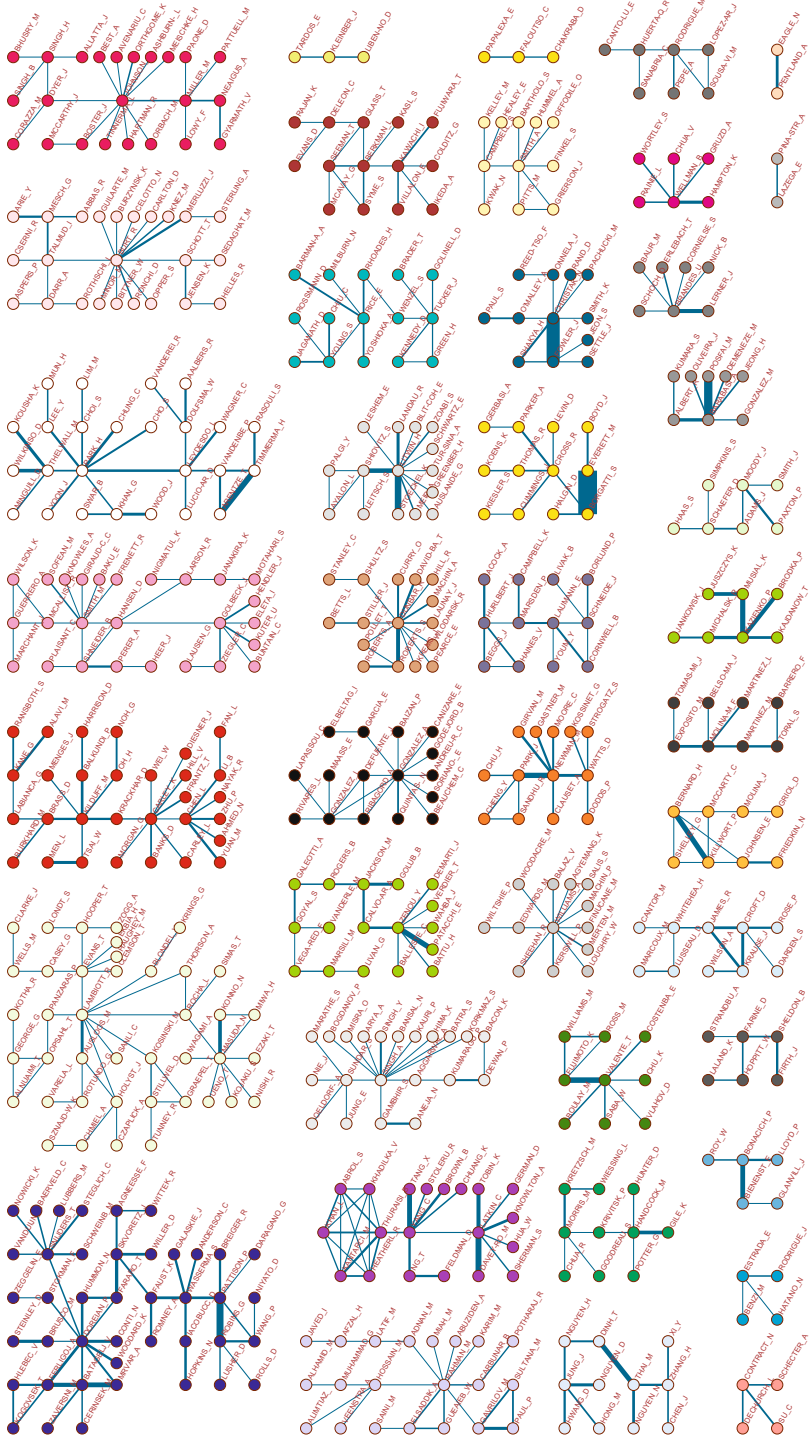


Fig. 11 Network Ct': general islands for selected authors

$$AK''[C, k] = \sum_{a \in C} AK''[a, k] \quad \text{and} \quad AJ''[C, j] = \sum_{a \in C} AJ''[a, j] \quad (14)$$

This approach can be used to get the data on keywords and journals for any interesting group of authors. Again, as there was no time and space to drill into each of the islands, we selected three general islands represented by *Wasserman* (42 authors), *Dunbar* (16 authors), and *Newman* (13 authors), and looked at the keywords and journals which are most typical for them.

Table 5 shows that in all three cases the most used keywords are trivial—*network* and *social*. The island represented by Wasserman has many commonly used keywords, such as *model, analysis, graph, structure, datum, structural, theory, method*. Two other islands contain fewer general scientific words, such as *analysis, structure, and theory* for the Dunbar et al. island, and *graph, structure, model, analysis, base, algorithm, datum, and theory* for the Newman et al. island. Other keywords provide some indication of the topics that are studied by the groups of authors, oriented toward methodological or substantive issues. The group represented by Wasserman shows its orientation toward methodological issues in SNA, by such keywords as *correction, equivalence, random, power, markov, evolution, statistical, dynamics, generalized, regression, exponential, blockmodel, logit, balance, p, cluster, logistic, bias, dynamic, blockmodeling*. The keywords for the group represented by Dunbar (*human, evolution, brain, primate, neocortex, evolutionary, cognitive, baboon, etc.*) show their orientation to social and evolutionary neuroscience, including constraints on social group size (and the social brain hypothesis), the structure and dynamics of contemporary human social networks, the cognitive and time constraints which limit the number of relationships, social cohesion, bonding, kinship and friendship (the associated keywords are marked in bold). The keywords for the group represented by Newman show the orientation of network scientists toward complex networks, highlighting the methodological developments (*complex, small, world, random, dynamics, scale, cluster*) and the substantive issues this group is working on (*collaboration, science, scientific, online, spread, food, disease, percolation*).

Table 6 shows the top journals used by the selected groups of authors. For the group represented by Wasserman, the top journal is *Social Networks*, followed, by a large margin, by the *The Journal of Mathematical Sociology, Psychometrika*, and *The Journal of Mathematical Psychology*. Among the other journals, there are many journals from sociology, social sciences and humanities, including *Network Science* journal; also there are such journals as *LNCS* (series of proceedings) and *Plos One* (megajournal). For the group represented by Dunbar, the top journal is *Human Nature—An Interdisciplinary Biosocial Perspective*, followed by *Evolution and Human Behavior, Philosophical Transactions of the Royal Society B: Biological Sciences, Social Networks*, and *Animal Behaviour*. Other journals are in the biological, cognitive, behavioral and social sciences; interdisciplinary general scientific journals *Plos One* and *PNAS* are also in the list. The top journal for the group represented by Newman is *Physical Review E*, followed, by a large margin, by *PNAS, Nature, Physical Review Letters, Social Networks*, and *Science*. Other journals are from a variety of disciplines from natural sciences, however, some journals from social sciences are also included (*American Journal of Sociology, Scientometrics*).

Table 5 Selected groups of authors and keywords (the keywords associated with the topics under study are marked in bold, the general keywords associated with SNA are marked in italics)

Wasserman et al.			Dunbar et al.		Newman et al.	
Rank	Value	Keyword	Value	Keyword	Value	Keyword
1	30.02	<i>network</i>	3.80	<i>social</i>	11.27	<i>network</i>
2	20.11	<i>social</i>	3.60	<i>network</i>	5.05	<i>social</i>
3	8.62	<i>model</i>	2.97	size	4.13	complex
4	7.36	<i>analysis</i>	1.41	human	3.80	world
5	6.01	<i>graph</i>	1.31	evolution	3.54	small
6	5.50	<i>structure</i>	1.17	brain	2.68	<i>graph</i>
7	3.19	<i>datum</i>	1.13	group	2.51	<i>structure</i>
8	3.03	<i>structural</i>	0.90	primate	2.43	web
9	3.00	correction	0.84	support	2.31	internet
10	2.96	exchange	0.79	perspective	2.24	<i>model</i>
11	2.80	equivalence	0.75	constraint	1.96	community
12	2.68	random	0.69	neocortex	1.82	degree
13	2.54	<i>theory</i>	0.69	relationship	1.46	random
14	2.53	power	0.61	communication	1.39	dynamics
15	2.51	markov	0.60	hypothesis	1.28	<i>analysis</i>
16	2.41	evolution	0.57	evolutionary	1.16	<i>base</i>
17	2.28	group	0.55	behavior	1.10	sequence
18	2.25	statistical	0.52	cooperation	1.03	percolation
19	2.19	<i>method</i>	0.51	cognitive	0.93	collaboration
20	2.18	dynamics	0.49	difference	0.92	online
21	1.88	generalized	0.47	kinship	0.90	centrality
22	1.82	journal	0.44	altruism	0.74	science
23	1.80	regression	0.43	bond	0.70	cluster
24	1.78	exponential	0.40	organization	0.69	scientific
25	1.78	blockmodel	0.40	emotional	0.67	<i>algorithm</i>
26	1.76	logit	0.38	female	0.67	access
27	1.73	balance	0.38	closeness	0.66	spread
28	1.73	p	0.36	<i>analysis</i>	0.65	control
29	1.68	<i>measure</i>	0.36	individual	0.64	solution
30	1.66	<i>algorithm</i>	0.35	baboon	0.64	scale
31	1.66	cluster	0.35	internet	0.62	<i>datum</i>
32	1.64	<i>approach</i>	0.35	dynamics	0.60	user
33	1.62	actor	0.34	<i>structure</i>	0.60	<i>theory</i>
34	1.59	logistic	0.34	personal	0.59	distribution
35	1.55	relation	0.33	volume	0.56	food
36	1.54	introduction	0.33	world	0.56	pattern
37	1.54	bias	0.32	community	0.56	use
38	1.51	dynamic	0.32	friendship	0.55	service
39	1.45	blockmodeling	0.31	<i>theory</i>	0.53	relationship
40	1.44	friendship	0.30	society	0.52	disease

Table 6 Selected groups of authors and journals

Wasserman et al.			Dunbar et al.		Newman et al.	
Rank	Value	Journal	Value	Journal	Value	Journal
1	122.60	Soc Networks	4.67	Hum Nature-Int Bios	35.02	Phys Rev E
2	25.18	J Math Sociol	4.17	Evol Hum Behav	8.62	P Natl Acad Sci
3	10.75	Psychometrika	3.33	Philos T R Soc B	5.25	Nature
4	8.08	J Math Psychol	2.95	Soc Networks	4.75	Phys Rev Lett
5	7.33	Sociol Method Res	2.75	Anim Behav	3.00	Soc Networks
6	7.10	J Classif	2.57	P Roy Soc B-Biol Sci	2.92	Science
7	7.08	Qual Quant	2.33	Biol Lett-Uk	2.83	Lect Notes Comput Sc
8	5.04	Sociol Methodol	2.25	J Theor Biol	2.75	Eur Phys J B
9	4.00	J Am Stat Assoc	2.20	Roy Soc Open Sci	2.50	Am J Sociol
10	4.00	Contemp Sociol	2.00	Pers Indiv Differ	2.00	J Stat Mech-Theory
11	3.90	Am Sociol Rev	2.00	Trends Cogn Sci	1.97	Plos One
12	3.59	Lect Notes Comput Sc	2.00	Behaviour	1.67	Siam Rev
13	3.58	Am J Sociol	1.70	Sci Rep-Uk	1.33	Reliab Eng Syst Sa
14	3.53	Scientometrics	1.67	Comput Hum Behav	1.33	Secur Commun Netw
15	3.40	Soc Forces	1.62	Plos One	1.25	Nat Commun
16	3.38	Sociol Method	1.58	Hum Nature	1.12	J Comput Sci-Neth
17	3.17	Brit J Math Stat Psy	1.57	Adapt Hum Behav Phys	1.08	Scientometrics
18	3.00	J Am Soc Inform Sci	1.53	P Natl Acad Sci	1.03	J Supercomput
19	2.98	Annu Rev Sociol	1.50	Brit J Psychol	1.00	Inform Process Ma
20	2.75	Soc Psychol Quart	1.25	Soc Cogn Affect Neur	1.00	Comm Com Inf Sc
21	2.50	Inform Process Manag	1.00	Int J Dev Disabil	1.00	Symmetry-Basel
22	2.19	Plos One	1.00	Group Dyn-Theor Res	1.00	Electron Libr
23	2.17	Netw Sci	1.00	Curr Dir Psychol Sci	1.00	Contemp Phys
24	2.00	Poetics	1.00	Brit J Dev Psychol	1.00	Comput Phys Commun
25	2.00	Sociol Rev	1.00	Soc Dev	1.00	J Consum Res
26	2.00	Soc Sci Res	1.00	J Community Appl Soc	1.00	J Stat Phys
27	2.00	Comput Stat Data An	1.00	Folia Primatol	1.00	Harvard Bus Rev
28	1.92	Methods Ser	1.00	Hist Hum Sci	1.00	Annu Rev Sociol
29	1.92	Organ Sci	1.00	Underst Complex Syst	1.00	Jpn J Polit Sci
30	1.75	Sociol Perspect	1.00	Orig Hum Behav	1.00	Phys Lett A

Conclusions

In this paper, we conducted the analysis of co-authorship networks to study the collaboration between authors in the field of SNA. The interest in this topic is motivated by the recent studies (Maltseva & Batagelj, 2019, 2020, 2021), which showed that SNA field is currently developing in the variety of disciplines, from the social and natural sciences, and is applied to different research topics and issues. The analysis of the most prolific authors and their collaboration with others reveals some important trends and sheds light on the current development of the field.

To observe the whole field, we used a comprehensive approach to data collection. The initial dataset was formed from publications matching the query “social network*”. As

some works related to a broader field of network analysis in general could have been overlooked, we extended the results of the original query with a saturation search of papers which were intensively cited, and included the works from the flagship journals indexed in WoS, and around 100 most prominent authors. The dataset covers not only the works of social scientists, but also influential papers published by physicists, biological scientists, information and computer scientists, etc. This resulted in 70,792 works with full descriptions and 93,011 authors. The two-mode authorship network linking works with their authors was used for the production of three types of collaboration networks, including their temporal dimensions. Analyzing the networks, we observed the publishing activity of authors in the field of SNA, looked at the general trends of collaboration, extracted the most prolific authors and their most collaborative groups.

The results of the analysis on the **macro-level** show that half of the works in our dataset are co-written by two or three authors, and only one fifth are single-authored papers. There is a strong trend toward collaboration in the field: even though the number of papers written by a pair of authors has remained quite stable from the 1980s—around 25%,—the number of single-authored papers dropped from about 70% to about 10% for the period under study. The number of papers written by three, four, five, or more authors has increased since 1980. Besides the general trend to more active collaboration, which is specific to modern science, this rise can also be due to the involvement of researchers from many disciplines, with writing cultures different from social sciences, where the field originally developed.

As in many cases of working with bibliographic data, our approach was challenged by the problem of entity resolution (synonyms and homonyms), especially by the problem of author name resolution (homographs), which is known in the literature as “multiple personalities” (Harzing, 2015). This problem would be simplified by the standardization of information stored in bibliographic databases (ORCID, DOI, ISSN, ISBN, etc.) and the identification of the authors with their personal IDs (ORCID); however, this information was not provided in our data source. While this does not create problems for the observation of general trends, on the level of the author and group identification the authors with Korean and Chinese names “blurred” some of the results, that is why, to our regret, we had to leave them out of this analysis, and did not use some measures, such as the number of co-authors, or P_s -cores, for the identification of the most prolific authors and their groups. We have to accept this as a limitation of the study.

Being particularly interested in the description of the SNA community, we identified and listed the most prolific authors with the largest input into the field, on the **micro-level** of analysis. This input can be measured as the total number of works that each author published, but also as the *fractional* (total) self-contribution of the authors to their own works and works written in collaboration; in both cases, the resulting table of the top-60 scholars lists well-known authors from the field. The comparison of the two measures shows that for some authors with a large number of works the total self-contribution is relatively low, which means that they achieve this large number of works through collaboration with other authors; however, some other prolific authors represent the opposite pattern, having a high value of self-contributions to the total number of their works. To measure the authors’ “willingness to collaborate”, from the two metrics, we calculated a special index K_a called “collaborativeness”, varying from 0 to 100%. The results show that the most prolific authors, in general, tend to collaborate, as the number of authors with K_a larger than 50% is large, while some authors are more self-sufficient. As an example, we looked at the temporal distributions for these metrics for the 10 most prolific authors—Latkin, Valente, Dunbar, Newman, Christakis, Doreian, Carley, Burt, Borgatti, and Snijders. We cannot make

general conclusions based on the observation of these distributions, but we can see that the authors' productivity can vary according to the periods, volumes and peaks of publishing activity. Interestingly, while some authors have been constant in the SNA field from the start of their careers, some authors are present only temporarily but make a considerable contribution to the field in the number of works, comparable to the SNA core researchers. The measures of collaborativeness can also be quite stable for some authors, but also vary through time, arising situationally, possibly depicting the biographical paths of the authors.

Measuring collaboration, we used two networks to obtain the groups of the most connected authors on the **meso-level** of analysis, based on the number of works co-authored, and their fractional contribution to the works written in collaboration (based on Newman's normalization, which does not take into account the collaboration of the author with himself or herself—considers only works with at least two authors). We found that the number of authors having a large number of co-authored papers (at least 15) is relatively small (114 authors), and in almost half of the cases these are pairs of authors intensively working together. The extraction of subgroups from the network, which takes into account the personal input of authors, through the Islands approach, showed that there are very many islands (more than 14,000) which have higher internal cohesion than the links to their neighbors, and most of them consist of a relatively small number of authors—78% of the nodes are in groups of not more than 5 researchers. The question arises of the approaches that should be used for the identification of the interesting subgroups for further investigation. The identification of the largest islands provided only partial insight, as some of the structures are star-like or cliques, and non-trivial structures do not necessarily represent important or well-known collaborators. We found it more useful to either look at the islands which contain the largest link weights, or simply choose the islands with authors of interest. And again, in many cases we found star-like structures of the groups of authors, connected around one or two well-known authors.

Concerning the disciplinary identity of the authors, the most prolific ones are from the *social* branch of network analysis. There are also representatives of *network science* discipline, computer science, epidemiological and health studies, and behavioral biology. This goes in line with the previous studies. Mainstream social network researchers were also identified in previous studies (Batagelj et al., 2014; Leydesdorff et al., 2008; Otte and Rousseau, 2002) the recent study of Batagelj et al. (2020) also contained physicists. In addition to this, the analysis of the disciplinary affiliations of other identified groups of collaborative authors revealed other topics and disciplines where network analysis is applied. Among the researchers from the social sciences—the traditional direction of SNA development—we extracted the subgroups working on substantive issues, such as personal social networks, social support, psychology, health and medical studies, risk and disease social networks, and online social networks and social media, social gerontology; as well as those working on methodological SNA issues, including ERGM, or UCINET. The groups of authors belonging to computer science and physics work on artificial intelligence, artificial neural networks, and complex networks. We also identified the groups of researchers applying network analysis to other fields, such as medicine, neuroscience, animal and human behavior, or urban studies. These are only the topics of the authors who appeared central in our analysis; other groups of authors may cover many other topics, which it was impossible to identify in this article. However, these disciplines can be seen as the ones where SNA is developing the fastest.

Using the general islands approach, which produces larger groups of authors, we noticed that several separate groups of social scientists merged into larger groups. There were no other groups from other disciplines, which merged into one cluster, however, some of them

significantly increased in size. This may mean that, in general, many groups of authors from the social sciences are connected to each other, while other authors working in separate topics within the same discipline, or other disciplines, tend to form separate groups.

Bibliometric analysis can provide a way to identify (the groups of) authors. In this article, we observed differences between the representatives of various streams of network analysis. We chose three subgroups of authors—represented by Wasserman, Dunbar, and Newman—and looked at the keywords and journals with which they are associated. Even though, in all cases, the most used keywords were trivial—*network* and *social*—we identified keywords describing the groups, showing their orientation toward methodological issues of SNA (Wasserman et al.), social and evolutionary neuroscience (Dunbar et al.) and complex networks analysis (Newman et al.). The analysis of the journals associated with these groups also showed the differences between them: the first group publishes papers in the journals from sociology, social sciences and humanities; second—in the biological, cognitive, behavioral, and social science journals; and third—mostly in natural science journals. This approach can be used for a description of other groups of collaborative authors.

The approach to the analysis of collaboration proposed in this study can be used in further studies aimed at the analysis of various scientific fields. In this study, the approach to temporal network analysis developed before (Batagelj & Maltseva, 2020) was applied to large bibliographic networks to see patterns of collaboration. As examples, we identified the 10 most prolific authors, and traced some of the metrics of their publishing productivity and activity through time. In future, such information can be used to cluster the authors with similar patterns and measure whole academic communities. The automatization of these processes is a task for the future development of the software support.

Acknowledgements All computations were performed using the program for large network analysis and visualization Pajek (De Nooy et al., 2018) and Python code based on the library Nets (Batagelj, 2020a). Visualizations of distributions and temporal quantities were produced in R. We appreciate the help of David Connolly (Academic Writing Center, HSE University, Moscow) with the proofreading of the article.

Funding This work is supported in part by the Slovenian Research Agency (research program P1-0294 and research projects J1-9187 and J5-2557), project COSTNET (COST Action CA15109), and prepared within the framework of the HSE University Basic Research Program.

References

- Batagelj, V. (2005). SN5: Network data for Vizards session at INSNA Sunbelt 2008. <http://vlado.fmf.uni-lj.si/pub/networks/data/WoS/SN5.zip>
- Batagelj, V. (2017). WoS2Pajek. Networks from Web of Science. Version 1.5. <http://vladowiki.fmf.uni-lj.si/doku.php?id=pajek:wos2pajek>
- Batagelj, V. (2020a). Nets. Python package for network analysis. <https://github.com/bavla/Nets>
- Batagelj, V. (2020b). On fractional approach to analysis of linked networks. *Scientometrics*, 123(2), 621–633. <https://doi.org/10.1007/s11192-020-03383-y>.
- Batagelj, V., & Cerinšek, M. (2013). On bibliographic networks. *Scientometrics*, 96(3), 845–864.
- Batagelj, V., Doreian, P., & V., Ferligoj, A., Kežar, N. (2014). *Understanding Large Temporal Networks and Spatial Networks: Exploration, Pattern Searching, Visualization and Network Evolution*. Wiley.
- Batagelj, V., Ferligoj, A., & Doreian, P. (2020). Bibliometric analysis of the network clustering literature. In P. Doreian, V. Batagelj, & A. Ferligoj (Eds.), *Advances in network clustering and blockmodeling*. Wiley.
- Batagelj, V., Ferligoj, A., & Squazzoni, F. (2017). The emergence of a field: A network analysis of research on peer review. *Scientometrics*, 113(1), 503–532. <https://doi.org/10.1007/s11192-017-2522-8>.

- Batagelj, V., & Maltseva, D. (2020). Temporal bibliographic networks. *Journal of Informetrics*, *14*(1), 1001006.
- Batagelj, V., & Praprotnik, S. (2016). An algebraic approach to temporal network analysis based on temporal quantities. *Social Network Analysis and Mining*, *6*(1), 1–22.
- Bonacich, P. (2004). The invasion of the physicists. *Social Networks*, *26*, 285–288.
- Borgatti, S. P., & Foster, P. C. (2003). The network paradigm in organizational research: A review and typology. *Journal of Management*, *29*(6), 991–1013.
- Brandes, U., & Pich, C. (2011). Explorative visualization of citation patterns in social network research. *Journal of Social Structure*, *12*(8), 1–19.
- Clarivate Analytics. (2020). <https://clarivate.com/webofsciencgroup/solutions/webofscience-platform/>
- Chen, C. (2005). Measuring the movement of a research paradigm. In *Visualization and Data Analysis 2005* (Vol. 5669, pp. 63–76). International Society for Optics and Photonics.
- De Nooy, W., Mrvar, A., & Batagelj, V. (2018). *Exploratory social network analysis with Pajek: Revised and expanded edition for updated software* (Vol. 46). Cambridge University Press.
- Freeman, L. C. (2004). *The development of social network analysis. A study in the sociology of science*. Empirical Press.
- Freeman, L. C. (2011). The development of social network analysis—with an emphasis on recent events. *The SAGE Handbook of Social Network Analysis*, *21*(3), 26–39.
- Garfield, E. (2004). Historiographic mapping of knowledge domains literature. *Journal of Information Science*, *30*(2), 119–145. <https://doi.org/10.1177/0165551504042802>.
- Gaufrriau, M., Larsen, P., Maye, I., Roulin-Perriard, A., & von Ins, M. (2007). Publication, cooperation and productivity measures in scientific research. *Scientometrics*, *73*(2), 175–214.
- Harzing, A. W. (2015). Health warning: Might contain multiple personalities—the problem of homonyms in Thomson Reuters Essential Science Indicators. *Scientometrics*, *105*(3), 2259–2270.
- Harzing, A. W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison. *Scientometrics*, *106*(2), 787–804.
- Hilbert, F., Barth, J., Gremm, J., Gros, D., Haiter, J., Henkel, M., et al. (2015). Coverage of academic citation databases compared with coverage of scientific social media: Personal publication lists as calibration parameters. *Online Information Review*, *39*(2), 255–264.
- Hidalgo, C. A. (2016). Disconnected, fragmented, or united? A trans-disciplinary review of network science. *Applied Network Science*, *1*(1), 6. <https://doi.org/10.1007/s41109-016-0010-3>.
- Holme, P., & Saramäki, J. (Eds.). (2019). *Temporal Network Theory*. Springer.
- Hummon, N. P., & Carley, K. (1993). Social networks as normal science. *Social Networks*, *15*(1), 71–106.
- Hummon, N. P., Doreian, P., & Freeman, L. C. (1990). Analyzing the structure of the centrality-productivity literature created between 1948 and 1979. *Science Communication*, *11*(4), 459–480.
- Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy*, *26*(1), 1–18.
- Kejžar, N., Černe, S. K., & Batagelj, V. (2010). Network analysis of works on clustering and classification from web of science. *Classification as a tool for research* (pp. 525–536). Springer.
- Laudel, G. (2002). What do we measure by co-authorships? *Research Evaluation*, *11*(1), 3–15.
- Lazer, D., Mergel, I., & Friedman, A. (2009). Co-citation of prominent social network articles in sociology: The evolving canon. *Connections*, *29*(1), 43–64.
- Leydesdorff, L., Schank, T., Scharnhorst, A., & De Nooy, W. (2008). Animating the development of Social Networks over time using a dynamic extension of multidimensional scaling. *El Profesional de Informacion*. <https://doi.org/10.3145/epi.2008.nov.04>.
- Lietz, H. (2009). *Diagnosing emerging science: The cases of the 'New Science of Networks' and Scientometrics*. Technology and Innovation Indicators and Knowledge Dynamics Visualization: PRIME-ENID Summer School on Science.
- Maltseva, D., & Batagelj, V. (2019). Social network analysis as a field of invasions: Bibliographic approach to study SNA development. *Scientometrics*, *121*(2), 1085–1128.
- Maltseva, D., & Batagelj, V. (2020). Towards a systematic description of the field using keywords analysis: Main topics in social networks. *Scientometrics*, *123*, 357–382.
- Maltseva, D., & Batagelj, V. (2021). Journals publishing social network analysis. *Scientometrics*, *126*(4), 3593–3620.
- Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & López-Cózar, E. D. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, *12*(4), 1160–1177.
- Newman, M. E. (2001). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E*, *64*(1), 016132.
- Newman, M. E. (2018). *Networks*. Oxford University Press.

- Otte, E., & Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. *Journal of information Science*, 28(6), 441–453.
- Shibata, N., Kajikawa, Y., & Matsushima, K. (2007). Topological analysis of citation networks to discover the future core articles. *Journal of the American Society for Information Science and Technology*, 58(6), 872–882. <https://doi.org/10.1002/asi.20529>.
- Shibata, N., Kajikawa, Y., Takeda, Y., & Matsushima, K. (2008). Detecting emerging research fronts based on topological measures in citation networks of scientific publications. *Technovation*, 28(11), 758–775.
- Shibata, N., Kajikawa, Y., Takeda, Y., & Matsushima, K. (2009). Comparative study on methods of detecting research fronts using different types of citation. *Journal of the American Society for Information Science and Technology*, 60(3), 571–580.
- Varga, A. V., & Nemeslaki, A. (2012). Do organizational network studies constitute a cohesive communicative field? Mapping the citation context of organizational network research. *Journal of Sociology and Social Anthropology*, 5(64), 349–364.
- Yan, E., Ding, Y. (2012). A framework of studying scholarly networks. In *Proceedings of the 17th international conference on science and technology indicators*. Science-Metrix and OST (pp. 917–926).