The structure and evolution of scientific collaboration in Serbian mathematical journals

**Miloš Savić · Mirjana Ivanović · Miloš Radovanović ·
Zoran Ognjanović · Aleksandar Pejović ·
Tatjana Jakšić Krüger**

Abstract Digital preservation of scientific papers enables their wider accessibility, but also provides a valuable source of information that can be used in a longitudinal scientometric study. The Electronic Library of the Mathematical Institute of the Serbian Academy of Sciences and Arts (eLib) digitizes the most prominent mathematical journals printed in Serbia. In this paper, we study a co-authorship network which represents collaborations among authors who published their papers in the eLib journals in an 80 year period (from 1932 to 2011). Such study enables us to identify patterns and long-term trends in scientific collaborations that are characteristic for a community which mainly consists of Serbian (Yugoslav) mathematicians. Analysis of connected components of the network reveals a topological diversity in the network structure: the network contains a large number of components whose sizes obey a power-law, the majority of components are isolated authors or small trivial components, but there is also a small number of relatively large, non-trivial components of connected authors. Our evolutionary analysis shows that the evolution of the network can be divided into six periods that are characterized by

M. Savić (✉) · M. Ivanović · M. Radovanović
Department of Mathematics and Informatics, Faculty of Sciences,
University of Novi Sad, Trg Dositeja Obradovića 4, 21000 Novi Sad, Serbia
e-mail: svc@dmi.uns.ac.rs

M. Ivanović
e-mail: mira@dmi.uns.ac.rs

M. Radovanović
e-mail: radacha@dmi.uns.ac.rs

Z. Ognjanović · A. Pejović · T. Jakšić Krüger
Mathematical Institute of the Serbian Academy of Sciences and Arts, Kneza Mihaila 36,
11001 Belgrade, Serbia
e-mail: zorano@mi.sanu.ac.rs

A. Pejović
e-mail: pejovica@mi.sanu.ac.rs

T. Jakšić Krüger
e-mail: tatjana@mi.sanu.ac.rs

different intensity and type of collaborative behavior among eLib authors. Analysis of author metrics shows that betweenness centrality is a better indicator of author productivity and long-term presence in the eLib journals than degree centrality. Moreover, the strength of correlation between productivity metrics and betweenness centrality increases as the network evolves suggesting that even more stronger correlation can be expected in the future.

Introduction

A well known feature of mathematical results is that they do not lose their significance over time. Digitization of mathematical journals in Serbia started in 1995 as a response to the increasing requirement for easier access to old issues of the journal *Publications de l'Institut Mathématique*. Later on, to enable access to the digitized material and to support preservation of old publications, in the year 2002 the Mathematical Institute of the Serbian Academy of Sciences and Arts decided to create an Internet database of freely accessible full-text mathematical journals - eLib (Mijajlović et al. 2010). ELib is defined as a web-orientated application which contains a collection of mathematical journals printed in Serbia and can be searched (both in English and Serbian), downloaded and printed.

The structure and evolution of collaboration among researchers can be investigated using so called *co-authorship networks*. The nodes in a co-authorship network represent researchers. Two researchers are connected by an undirected link if they authored at least one paper together, with or without other co-authors. The first studies of co-authorship networks in the field of mathematics (Grossman and Ion 1995; Batagelj and Mrvar 2000) have been deeply influenced by the existence of one prominent mathematician, Paul Erdős (1913–1996). His unique work ethic and lifestyle led to the publication of over 1500 papers with a great number of different co-authors, and inspired the concept of the Erdős number. More general analysis of mathematics collaboration networks is performed by Grossman (2002a, b) who examines statistical properties of the network derived from *Mathematical Reviews* (MR), and Brunson et al. (2012), who study the evolution of the MR network. Studies that focus on other research communities include investigations of co-authorship networks that emerged from different bibliographic databases covering different scientific disciplines (Newman 2001a, 2004b; Barabási et al. 2002; Bettencourt et al. 2009), multiple digital libraries covering computer science conferences (Liu et al. 2005), collaboration in the field of information systems (Cunningham and Dillon 1997), and information retrieval (Ding et al. 1998), collaboration among researchers from individual countries such as Slovenia (Perc 2010), and even individual publication venues like *Scientometrics* (Hou et al. 2008), and PNAS journals (Börner et al. 2004) or the SIGMOD conference (Nascimento et al. 2003).

Co-authorship networks can be viewed as ordinary social networks restricted to people doing science: links in a co-authorship network denote temporal and collegial

relationships, and imply a strong academic bond. Information available in co-authorship networks aggregated with other sources of information like keywords can be used to develop models for ranking most influential authors in a database (Mimno and McCallum 2007; Gollapalli et al. 2011), to automatically determine the most appropriate reviewers for a manuscript (Rodriguez and Bollen 2008; Kalmukov 2012), or even to predict future research collaborations (Liben-Nowell and Kleinberg 2003). Co-authorship networks are quite different from so called *citation networks* (Garfield and Sher 1963; Solla Price 1965) which are another type of complex networks related to scientometrics. Namely, citation networks reflect the structure of our scientific knowledge, while co-authorship networks depict the structure of academic societies.

In this paper we study the co-authorship network extracted from the bibliographical records contained in the electronic library of the Mathematical Institute of the Serbian Academy of Sciences and Arts (from this point on we use the term "eLib" to denote this particular digital library). The nature of the bibliographic data enables us to investigate the structure of scientific collaborations characteristic to authors that publish papers in the most prestigious Serbian mathematical journals, where the majority of present authors are Serbian mathematicians. The bibliographic records also cover a wide time range: the first research article indexed in eLib was published in 1932. Thus, we investigated the evolution of the eLib co-authorship network in an 80 year period, from 1932 to 2011, with yearly resolution in order to observe general trends in the evolution of collaborations among authors from the eLib community.

The rest of the paper is structured as follows. Section 2 explains the construction of the eLib co-authorship network. The next section presents the methodology that is used in the analysis of the network. The obtained results are presented and discussed in Section 4. Finally, the last section concludes the paper.

Construction of the eLib co-authorship network

This study is based on bibliographic data contained in eLib digital library. ELib digitizes and indexes articles published in the following Serbian journals:

1. "Publications de l'Institut Mathématique" (PIM), published by the Mathematical Institute of the SASA[1] since 1932,
2. "Matematički Vesnik" (MV), published by the Mathematical Society of Serbia since 1949,
3. "Zbornik Radova" (ZR), published by the Mathematical Institute of the SASA since 1951,
4. "Publications of Department of Astronomy" (PDA), published by the Faculty of Mathematics, University of Belgrade, since 1969,
5. "Nastava Matematike" (NM), published by the Mathematical Society of Serbia since 1992,
6. "The Teaching of Mathematics" (TTM), published by the Mathematical Society of Serbia since 1998,
7. "Visual Mathematics" (VM), published by the Mathematical Institute of the SASA since 1999,

---

[1] Serbian Academy of Sciences and Arts.

8. "Kragujevac Journal of Mathematics" (KJM), published by the Faculty of Sciences, University of Kragujevac, since 2000,

9. "Bulletin, Classe des Sciences Mathématiques et Naturelles, Sciences mathématiques" (Bulletin), published by the SASA since 2001,

10. "Review of the National Center for Digitization" (RNCD), published by Faculty of Mathematics, University of Belgrade, since 2002,

11. "Computer Science and Information Systems" (ComSIS), published by the ComSIS Consortium[2] since 2004.

The eLib database was designed primarily with an e-library web application in mind, i.e. it was designed as a backend part of an e-library application. This means that its design was optimized for the needs of the e-library application and as such it is not particularly suited for extensive data analysis. Because of the above, we decided to export a subset of the database into a textual file and use it for subsequent analyses. In order to achieve this we needed to determine what data was required for the analysis that we were going to perform. Additionally, in our analysis we wanted to distinguish between male and female authors. Since gender data was not readily available, we needed to determine the gender of each author based on their name. Due to the existence of initials instead of full first names (especially in the earliest issues of the journals), we found it quite difficult to do this algorithmically and decided to do it by hand. We started by assuming that all the authors are male[3] and then we considered each name in the data one by one. Whenever we determined with certainty that the author was female, we would mark that in the database record. Otherwise, we would leave it unchanged. Sometimes we could conclude the gender of an author solely from her/his name. In more difficult cases we used the Internet, or solicited the assistance of a few retired professors, older colleagues and the corresponding editors who actually knew the authors of some papers from 70–80 years ago. Finally, this procedure ensured that we got a good lower bound of the number of female authors in the journals that we have analyzed.

When we completed the data requirements, we developed a procedure that denormalizes data from the e-library database and exports it in a CSV file. We ended up with one record per paper that contains the following information: ID (identifier of paper in the e-library database), language the paper is written in, paper title, list of keywords, MSC classification, journal in which the paper was published, year of publication, number of pages, first name, last name and gender of each author.

The construction of the eLib co-authorship network is done in five phases:

1. First analysis of author names: this phase resulted in the identification of authors and the formation of a lookup table which is used to correct author names.
2. Construction of the inverted index which maps authors to papers they published.
3. Construction of the initial co-authorship network from the inverted index.
4. Second analysis of author names: the name analysis procedure from step one is performed on each connected component independently in order to detect potential name lookup entries that were not detected in the first analysis of author names.
5. Steps 2 and 3 are repeated in order to obtain the final co-authorship network.

---

[2] The ComSIS Consortium is a group of leading scientific institutions from universities in Serbia including the Serbian Academy of Sciences and Arts, who jointly publish the ComSIS journal.

[3] Due to our familiarity with the Serbian mathematics scene it was quite reasonable to assume that the majority of eLib authors are male.

The analysis of author names is conducted to identify authors in the exported data. This step is especially important, because the nodes in co-authorship networks are identified by researchers' names, and there might be spelling errors and other inconsistencies in bibliographic records. The names in exported data can be divided into two categories: full names (provided both full first name and last name) and short names (first name is reduced to first letter(s)). Exported eLib bibliographic data contains 8,842 name instances in total, where 5,192 names (58.72 %) are full names, while 3,650 names (41.28 %) are short names. Of course, there is a source of error in distinguishing different researchers having the same name, regardless of the name being full or short. As observed by Barabási et al. (2002), who identified researchers by last name and initials of first name, this error may affect results of analysis of co-authorship graphs mainly for scientists of Chinese and Japanese descent. Newman (2001a), who also identified researchers by last name and initials of first name, showed that the errors in results caused by errors in the identification of unique authors are of the order of a few percent. Our results are also affected by errors in the identification of unique authors. However, in order to reduce them to the minimum, we conducted an analysis of names that appears in our data.

Our analysis of author names is based on the usage of three string similarity measures provided by the LingPipe[4] Java library: Jaccard, Jaro-Winkler and TF-IDF. String similarity measures are computed independently for (1) name pairs from the set of unique full names, (2) name pairs from the set of unique short names, and (3) the Cartesian product of unique full and short names, where a full name can be uniquely reduced to an appropriate short name, i.e. there are no two full names in the data that can be reduced to the same short name. Computed string similarity measured are sorted, cut at the threshold 0.6 for each used string similarity measure, and inspected manually in order to form the name lookup table that is used to correct author names. In this process, when necessary, we also checked information about authors contained in publications written by authors from selected name pairs.

In the second stage of the co-authorship network construction the inverted index is formed. This structure, denoted by I, maps the name of an author (after lookup correction) to the set of papers he or she (co-)authored. The set of keys of the inverted index corresponds to the set of nodes in the co-authorship network. Two authors A and B, where A and B are two keys from the inverted index, are connected by the undirected link $A \leftrightarrow B$ if and only if $I(A) \cap I(B) \neq \varnothing$, where $I(A)$ denotes set of papers published by A. The cardinality of $I(A) \cap I(B)$ (the number of papers A and B co-authored together) is the weight of link $A \leftrightarrow B$.

The analysis of author names is repeated after the first construction of the co-authorship network, but this time the name pairs are formed from connected components of the network. This means that string distances are computed for researchers contained in the same connected component. This step was motivated by the following observation: if two similar names represent the same author then there is a high probability that they have at least one co-author in common, especially in the case when we are looking for matches between short and full names.

Analysis of the eLib co-authorship network

The publication dynamics of the eLib journals is investigated by measuring the number of papers at a yearly level. The construction of the eLib co-authorship network enables us to

---

[4] http://alias-i.com/lingpipe/.

examine other context-relevant static and dynamic aspects of the eLib community: the number of authors (where we distinguish between male and female authors), the fraction of "returning" authors, the average number of authors per paper, the fraction of single-authored papers, the distribution of the number of papers per author, and the distribution of the number of authors per paper.

Definition 1 (*Returning*, *old author*; *new author*) An author is called returning or old if he or she already published paper in one of the eLib journals. An author that publishes a paper for the first time in an eLib journal is called a new author.

Definition 2 (*Author timespan*) The timespan for author A is defined as the number of years that passed from the publication of A's first article to the publication of A's last article in eLib journals. If A published exactly one paper in the eLib journals then A has timespan equal to one.

A more recent resurgence of interest in networks of scientists was sparked by the observation of power-law degree distributions in various types of real-world networks (Barabási and Albert 1999) including co-authorship networks (Newman 2001a, b, 2004a, b; Barabási et al. 2002). On the other hand, power-laws in statistical properties of scientific networks like the number of papers per author were observed (by hand) way back in the early 20th century (Lotka 1926) and confirmed by the subsequent (computerized) studies (Voos 1974; Pao 1986; Newman 2004b).

Definition 3 (*Power-law*) A quantity k obeys a power-law if it is drawn from a probability distribution $P(k) \propto k^{-\gamma}$, where γ is a constant parameter of the distribution.

In order to determine whether a distribution of highly dispersive empirical data obeys a power law we examined the complementary cumulative distribution function $CCD(k) = \sum_{i=k}^{\infty} P(i)$ instead of $P(k)$, because CCD reduces noise that may appear in the tail of the distribution where large but rare events occur (Newman 2005). If $P(k)$ obeys a power-law, that is $P(k) \propto k^{-\gamma}$, then $CCD(k) \propto k^{-(\gamma-1)}$. Therefore, the plot of $CCD(k)$ on log-log scales appears as a straight line. This makes power-laws particularly easy to spot by plotting the complementary cumulative distribution function on logarithmic scales. To determine the power-law scaling exponent (γ) we used the non-linear least square fitting procedure based on the Levenberg–Marquardt algorithm implemented in MicroCal Origin[5] data analysis software.

The analysis of structure of scientific collaborations in the eLib journals is based on standard methods and metrics used in analysis of social networks. Connected component analysis is conducted to determine properties of connected components contained in the network. A connected component of an undirected network is a set of mutually reachable nodes, i.e. there is a path connecting each two nodes in the component. We distinguish between two types of components in a co-authorship network: non-trivial and trivial components.

Definition 4 (*Trivial component*) A component of a co-authorship network is considered trivial if it is a complete sub-graph of the network (each two nodes in the component are directly connected), and the weight of each link is equal to one.

Trivial components represent collaborations established by publication of exactly one paper. If there is a group of authors that published exactly one paper together, and if this

---

[5] http://www.originlab.com/.

paper is the only published paper for each member of the group, then the members of the group form a trivial component.

Nodes (authors) in the co-authorship network are quantified by the following metrics that are widely used in analysis of social networks:

1. *Degree centrality*—Degree centrality of node *A* is the number of other nodes *A* is directly connected with. In other words, degree centrality of *A* is the number of other authors with whom *A* collaborated.
2. *Betweenness centrality*—Betweenness centrality (BC) of node *A* is equal to the number of shortest paths from all nodes to all other nodes that pass through *A*. BC quantifies the importance of a node within a network. Nodes with higher BC are more important since they play the role of a broker which connects other nodes. Due to their vital role to the overall connectedness of a network or individual connected components, nodes with high BC are considered as the most influential actors in a group, organization or community that is represented by the network. Betweenness centrality can also be computed for links. BC of link *L* is the number of shortest paths from all nodes to all other nodes that pass through link *L*.
3. *Small-world coefficient*—Small world coefficient of node *A* is the average length of shortest paths connecting *A* and all reachable nodes (authors).
4. *Clustering coefficient*—Clustering coefficient (CC) of author *A* in a co-authorship graph denotes the probability that two co-authors of *A* are co-authors themselves. In other words, CC of *A* is the density of the sub-graph that is induced by *A* and authors to which *A* is directly connected.

Definition 5 (*Isolated author*) An author is called isolated if his degree centrality is equal to zero. Isolated nodes in the eLib co-authorship network represent authors who have not collaborated with other authors from the eLib community by publishing joint papers in the eLib journals.

To evaluate the productivity of researchers Lindsey (1980) suggests the following methods: normal count, fractional (adjusted) count, and straight count. Normal count gives every author one credit, straight count assigns all the credit to the first author only, while fractional count assigns credit equal to $1/n$ to each of the n co-authors. In this study, we use the normal counting method for measuring author's productivity. Additionally, timespan (see Definition 2) is recorded for each author, since it is a metric of long-term presence in the eLib journals. Correlations between author metrics are investigated by the computation of Spearman's rank correlation coefficient.

Small-world and clustering coefficients for components are calculated by averaging the values of mentioned metrics for all authors from the components. The clustering coefficient of a component quantifies its cohesiveness, i.e. how close the component is to being a clique (complete graph). The number of publications for a component is the total number of publications written by authors from the component.

Definition 6 (*Component timespan*) The timespan of a component is the number of years that passed from the creation of the component to the last event which changed the structure of the component.

Each component is created by publishing a paper that is written by one or more new authors. Therefore, the year when the paper was published determines the time of creation of the component. At the time of creation each component is either an isolated node or trivial component. There are two types of evolutionary events that change the structure of

the component: inclusion of a new author into the component and renewal of collaborations among authors from the component. Inclusion of new authors is achieved by the establishment of a collaboration between an old author from the component and an author that published a paper in the eLib journals for the first time. This event causes the creation of one new node and one new link in the component. Renewal of collaborations is manifested by an increase of link weights.

To investigate the evolution of the eLib co-authorship network we construct time-ordered snapshots of the network, i.e. the sequence of networks

$$S = \langle N_y \rangle, 1932 \leq y \leq 2011,$$

where N_y denotes the snapshot of the network in the year y. The set of nodes of N_y contains all authors that published at least one article before or during y. A link between authors A and B is present in N_y if A and B established collaboration before or during y. Accordingly, the weight of link $A \leftrightarrow B$ is the number of papers A and B authored together before or during y. Since co-authorship networks evolve by adding nodes and edges, the sequence S satisfies the following property: network N_y is a sub-graph of network N_w if $y < w$. The evolution of a quantifiable property P of the eLib co-authorship network is investigated by the examination of the numerical sequence $P(S) = \langle P(N_y) \rangle$.

The main property of trivial components is that they do not evolve. Evolutionary properties of non-trivial components are investigated by constructing sequence S from the full eLib co-authorship network (network N_{2011}) after removing isolated nodes and trivial connected components. This means that isolated nodes and trivial connected components exist in a network within S only if they lose mentioned properties in one of subsequent network snapshots.

In the evolutionary analysis we also distinguish three types of collaborations: collaborations between old (returning) authors, old and new authors, and new authors. In the computation of the number of collaborations link weights have to be taken into account. Let A and B be two authors connected by link $A \leftrightarrow B$ of weight w in network snapshot N_y. The following cases are possible:

1. Authors A and B do not exist in network N_{y-1}. Then link $A \leftrightarrow B$ in year y denotes one collaboration between A and B as new authors and $w - 1$ collaborations between A and B as old authors.
2. A exists and B does not exist in N_{y-1}. Then link $A \leftrightarrow B$ in year y denotes one collaboration between old author A and new author B, and $w - 1$ collaborations between A and B as old authors.
3. Both A and B exist in N_{y-1} and they are connected by a link of weight z. In this case link $A \leftrightarrow B$ in year y denotes $w - z$ collaborations between A and B as old authors.

Results and discussion

Basic results

In total 6,480 research papers were published in the eLib journals from 1932 to 2011. Figure 1 shows the publication dynamics of eLib, i.e. the number of papers published per year. It can be noticed that there are several periods in the evolution of eLib where the number of papers per year exhibits an increasing trend, as well as several periods when it

shows a decreasing trend. Also, it can be observed that there was no scientific production in the eLib journals during and immediately after the Second World War (1939–1946). In the first five years of eLib, the number of published articles had a relatively stable evolution which means that there was no drastic increase or decrease in the number of published papers. The first long-term growth trend in the number of publications appeared during the Informbiro period (1947–1953) which is characterized by the conflict between Yugoslavia and the Soviet Union. In that time two new Yugoslav mathematical journals were founded, and the number of publications increased from 15 in 1947 to 81 in 1953. After the Informbiro period the scientific production exhibited a general decreasing trend which ended by 1963. After 1963 the number of published papers started to increase and this trend ended in 1979. In the mentioned year the highest number of published papers before Yugoslav breakup is recorded (164 published papers).

After the death of Yugoslav president Josip Broz Tito in 1980, the economic crisis and national tensions in Yugoslavia started to emerge, leading to the Yugoslav breakup in 1991 and ethnic wars in the period from 1991 to 1995. These events evidently affected scientific production of the eLib community: Fig. 1 shows that the largest continual decrease in the number of papers per year occured in the period from 1980 to 1996. Especially in the war period (1991–1995) an extremely low degree of scientific production can be observed. In the study of mathematics scientific production from our neighboring country, Dravec Braun (2012) correlates the impact of the Serbo-Croatian war and establishment of universities and institutes with the Croatian mathematics scene, highlighting the two periods of stability and development before and after 1993. Although the war activities were not present on the territory of Serbia, the country was faced with the international sanctions which caused hyperinflation and brain drain. In that period the government's funds for education and science were drastically reduced, and researchers, mostly worried how to survive in such hard-living conditions, were demotivated for scientific work and publishing.

After the Dayton peace agreement (end of 1995), relative stabilization of the political situation in former Yugoslav republics caused a growth in the number of publications. However, the largest discontinuity in the number of published papers per year occured after the so-called "October revolution" (downfall of Slobodan Milošević government by the end of 2000), when the country entered a transitional period towards free-market economy and started to open to the rest of the world. Due to the new rules set by the Serbian ministry of science which emphasized the number of publications in journals as the main criterion for evaluation of scientific work, after 2001 the number of articles in the eLib journals per year is significantly higher than in previous years. Additionally, in the last years four new journals indexed by eLib are founded.

The total number of authors that published papers in the eLib journals during the examined period is 3,597, where 3,147 (87.49 %) authors are male and 450 (12.51 %) authors are female. Figure 2a shows the number of authors per year. It can be observed that the evolution of the number of authors per year has similar shape to the evolution of the number of papers per year. Pearson's correlation coefficient between these two variables is 0.929, while Spearman's correlation coefficient is 0.981. Additionally, in each year the number of male authors is significantly higher than the number of female authors (Fig. 2b). The smallest ratio between male and female authors was in 2007 (193 male and 69 female authors), while the largest ratio (excluding years when there were no female authors – period from 1933 to 1951, 1954, and 1962) was in 1968 (75 male and 1 female author).

The fraction of returning authors per year is shown in Fig. 3. It can be seen that two periods considering returning authors can be distinguished. After 1998 the fraction of
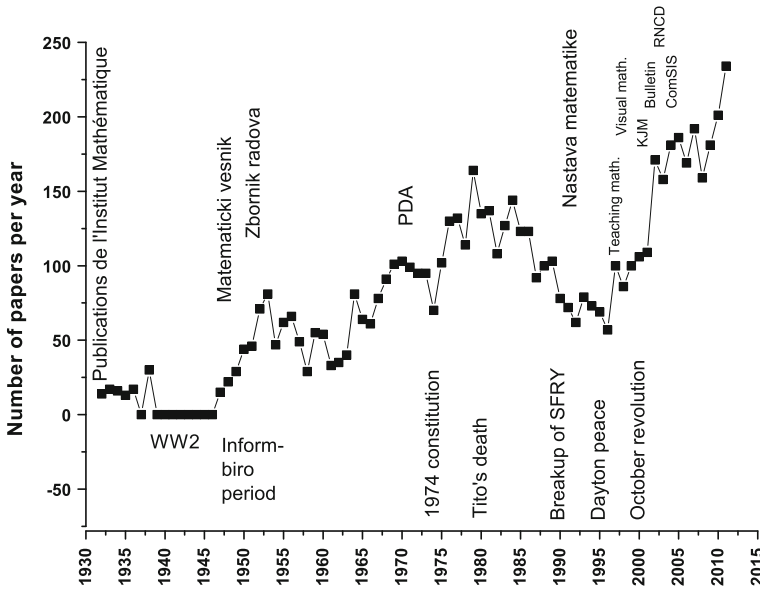
Fig. 1 The number of papers published in the eLib journals per year. Above the line are shown the names (or abbreviations) of journals in the time they were founded, while important events in Yugoslav/Serbian history are positioned below the line
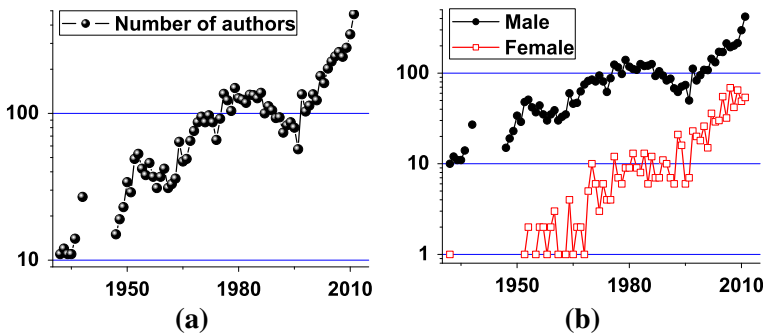

Fig. 2 The number of authors (**a**) and the number of male and female authors (**b**) per year

returning authors is always smaller than 0.5, which means that the majority of authors are new authors. In contrast, before 1998, in the majority of years, the majority of authors were returning authors. The notable exception is the year 1949 when the fraction of returning authors is the lowest during the whole eLib evolution. In that year 74 % of authors were new authors.

The majority of articles in the eLib journals in the investigated timeframe are single-authored papers: 4836 papers (74.63 % of the total number of papers) are written by exactly one author. This situation is not surprising for mathematical journals, since researchers in mathematics and humanities usually engage in solitary work, while laboratory scientists tend to write articles with many co-authors. Figure 4 shows the evolution

Fig. 3 The fraction of returning authors per year
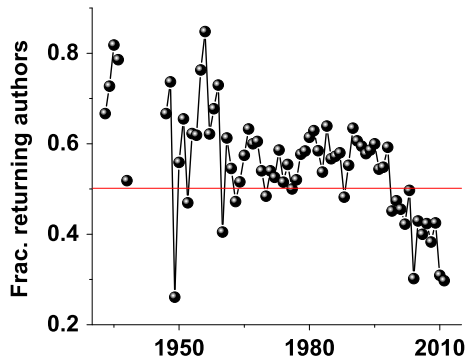
of the average number of authors per paper and the fraction of single-authored papers. It can be seen that the average number of authors per paper increases, while the fraction of single-authored papers decreases as eLib evolves. A similar evolutionary trend was also observed for articles indexed in *Mathematical Reviews* in the period from 1940 to 2000 (Grossman 2002a). As can be seen in Fig. 4a, the average number of authors per paper was slowly increasing from 1 to only 1.56 in the period from 1932 to 2005. However, in the last years (2005–2011) the average number of authors per paper has been growing significantly faster than in previous years, reaching 2.29 authors per paper in 2011. One of the factors which caused such fast growth is the foundation of new journals, RNCD and ComSIS, whose scope is not purely mathematical, but oriented to applications of mathematics and computer science, where the number of authors per paper is generally higher compared to pure mathematical research. Naturally, as the average number of authors per paper increases the fraction of single-authored papers decreases. More than half of the papers per year in the period 1932–2005 are single-authored papers, and only in the last years of eLib evolution the majority of papers were written by two or more authors.

Figure 5a shows the complementary cumulative distribution of the number of papers (k) per author plotted on log-log scales. It can be seen that the distribution closely follows a power-law with a faster decay (so called "cutoff" or truncated power-law) for $k > 25$. The same phenomena was also observed for the distribution of the number of papers per author in Los Alamos electronic preprint archive (Newman 2001a). The power-law nature of the distribution implies that the majority of authors published a small number of papers that is close to the average value. On the other hand, there is a small, but statistically significant, fraction of authors whose production is extremely higher compared to the average eLib author. For example, the most productive author present in the eLib journals is Ivan Gutman who published 71 papers in total, which is drastically higher than the average number of papers per author (the average number of papers per author is 2.458). The emergence of power-laws in empirical data can be explained by the principle of cumulative advantage (also known as "rich get richer" or preferential attachment principle). When applied to the distribution of the number of papers per author, the principle of cumulative advantage denotes that the probability that author A will publish a paper in the future is proportional to the number of papers A already published. In other words, the principle states that author A who at some point in time published more papers than author B has higher probability to publish a paper in the future than B. Observed cutoff in the distribution also has a natural, evolutionary explanation. Cutoffs in power-law distributions appear when time or capacity constraints are incorporated into the principle of cumulative
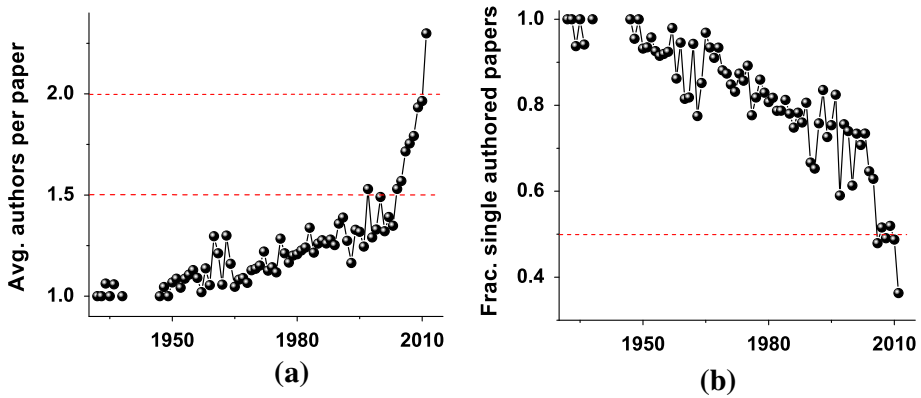
Fig. 4 The evolution of the average number of authors per paper (**a**), and the fraction of single authored papers (**b**)

advantage (Amaral et al. 2000). Even the most productive authors after some time stop publishing papers (due to retirement or death) thus introducing time constraints to the principle of cumulative advantage which governs the inequalities in the number of published papers per author.

In contrast to the heavy-tailed distribution of the number of papers per author, the distribution of the number of authors per paper (Fig. 5b) possesses a characteristic scale. This means that there are no papers with an extremely large number of authors: the maximal number of authors per paper is equal to ten, [6] while the average number and standard deviation of authors per paper are equal to 1.36 and 0.756, respectively. The majority of eLib papers are written by exactly one author (74.63 %), 17.91 % by two authors, 5.07 % by three authors, and only 2.37 % of the total number of publications have more than three authors.

Structure of the eLib co-authorship network

The co-authorship network formed from eLib bibliographic records contains 3,597 nodes (authors) and 2,766 links (collaborations). Basic quantities describing structural properties of the eLib co-authorship network are summarized in Table 1. In contrast to the co-authorship networks analyzed by Newman (2001a); Grossman (2002b); Barabási et al. (2002); Nascimento et al. (2003); Liu et al. (2005); Bettencourt et al. (2009), and Perc (2010), where the existence of a giant connected component is observed, the eLib co-authorship network is extremely fragmented: it contains 625 connected components neither of which is a giant connected component. The network exhibits small average shortest path lengths and a drastically larger clustering coefficient than the clustering coefficient of comparable Erdős-Renyi random graph. The clustering coefficient of the random graph with $N = 3579$ nodes and $L = 2766$ links is equal to $2N/L(L-1) = 0.00094$, while the clustering coefficient of the eLib co-authorship network is 0.44. This means that the eLib co-authorship network exhibits the Watts-Strogatz small-world property (Watts and Strogatz 1998; Newman 2001b).

---

[6] The article "Serbian Virtual Observatory" published in "Review of the National Center for Digitization" in 2009 is the article with the highest number of authors per paper.
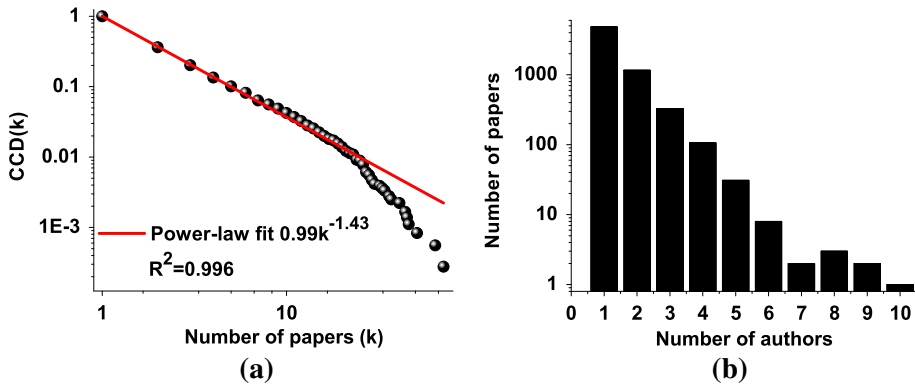
Fig. 5 Complementary cumulative distribution of the number of papers per author (**a**), and the distribution of the number of authors per paper (**b**)

Table 1 also provides basic structural quantities of co-authorship networks that are restricted to individual journals indexed by eLib. It can be seen that journal co-authorship networks mostly have similar structural characteristics as the eLib network: they are sparse, fragmented (a large number of connected components), do not contain a giant connected component, have a significant number of isolated nodes, and exhibit the small-world property. The number of papers published in an eLib journal strongly correlates with the number of authors and the number of components present in the journal: Pearson's correlation coefficient (PCC) between the number of papers and the number of authors is 0.959 (the value of Spearman's correlation coefficient is 0.718), while the PCC between the number of papers and the number of components is 0.909 (Spearman's correlation coefficient is 0.711).

Bibliometric indicators such as impact factor and h-index are commonly used to evaluate and compare scientific impact of journals (Ivanović et al. 2012). To the contrary, structural properties of journal co-authorship networks enable us to observe differences between journals that are related to the collaborative behaviour of authors rather than the scientific impact of their work. The smallest fraction of isolated nodes, only 3 % of the total number of authors, and the highest clustering coefficient is exhibited by the co-authorship network representing collaborations in the "Computer Science and Information Systems" (ComSIS) journal. This means that this journal mostly publishes papers with two or more authors, and has the most cohesive community of authors compared to other journals. Additionally, this journal has the highest link-node ratio, as well as the ratio between the number of links and the number of non-isolated nodes, which means that ComSIS authors established higher intensity of collaborations compared to other journals. The highest degree of collaborative behaviour exhibited by ComSIS authors can be explained by the fact that ComSIS is the only computer science journal indexed by eLib. Generally speaking, research in computer science, due to its experimental and applicative (industrial) component, mostly requires effort of a group of people, and consequently has higher collaborative potential compared to research in mathematics. The largest fraction of isolated nodes is in the co-authorship network of "Nastava matematike", where more than 60 % of the total number of authors are isolated. This means that this journal publishes mostly single-authored papers. The largest male–female ratio is in "Matematički vesnik", every fifteenth author present in this journal is female. In contrast, "Review of the National

Table 1 Basic structural parameters of the eLib co-authorship network and co-authorship networks resticted to individual journals

Net	#P	#N	#L	I	MFR	#C	LC	SW	CC
ELIB	6,480	3,597	2,766	0.33	6.99	625	0.06	2.117	0.44
PIM	2,150	1,147	563	0.42	8.10	202	0.05	1.721	0.25
PDA	122	48	21	0.47	8.60	7	0.12	1.421	0.30
MV	2,264	1,255	576	0.46	15.51	225	0.01	1.355	0.27
ComSIS	202	520	766	0.03	9.40	124	0.06	1.235	0.81
TTM	125	102	66	0.35	5.00	19	0.11	1.426	0.45
Bulletin	84	84	87	0.15	4.25	13	0.38	1.843	0.59
KJM	256	294	226	0.26	4.34	66	0.06	1.266	0.49
NM	352	195	58	0.63	2.09	24	0.04	1.293	0.26
ZR	327	172	73	0.54	5.61	24	0.05	1.378	0.46
RNCD	338	236	249	0.24	1.62	44	0.06	1.304	0.65
VM	260	212	182	0.36	6.85	43	0.08	1.224	0.39

#P the number of papers, *#N* the number of nodes (authors), *#L* the number of links, *I* the fraction of isolated authors, *MFR* male–female ratio, *#C* the number of connected components (isolated nodes are excluded), *LC* the relative size of the largest connected component, *SW* small-world coefficient, *CC* clustering coefficient

Center for Digitization" has male–female ratio 1.69 which means that this journal has the smallest gender gap among eLib journals.

As already mentioned, the eLib co-authorship network contains a large number of connected components, neither of them being giant. Figure 6a shows the complementary cumulative distribution of component sizes. It can be seen that the distribution can be very well approximated by the power-law with the scaling exponent $\gamma = 2.69$ (the coefficient of determination is $R^2 = 0.99$). This means that the majority of components are small-size components, but there are also connected components whose size is much larger than the average component size. The power-law scaling also appears in the distribution of the number of papers written by authors from the same component (Fig. 6b, $\gamma = 2.02$, $R^2 = 0.99$). Let A and B denote two connected components, where component A is larger than component B. The principle of cumulative advantage in the case of these two distributions suggests the following:

1. There is a higher probability that the eLib community will be expanded with an author who knows or collaborates with authors from component A than with those that are contained in the smaller component B. The mentioned probability is proportional to component size.
2. It is more probable that a newly published eLib paper, written by at least one returning author, will be written by authors from component A than authors from component B. Again, the probability is proportional to the number of papers written by authors from the connected component.

The eLib co-authorship network contains nearly the same number of trivial and non-trivial components: 319 components are trivial (51.04 %), while 306 of them are non-trivial. There are three largest trivial components, where each of them contains six authors. On the other side, there are 126 smallest non-trivial components. Those non-trivial components are groups of two authors which published more than one paper together in the

eLib journals, but have not collaborated with any other eLib author. The structural characteristics of ten largest connected components are shown in Table 2. The largest connected component encompasses 249 authors, which is only 6 % of the total number of authors, or 10.49 % of the total number of non-isolated authors. The number of papers published by authors from the largest component is 997, which is 15.38 % of the total number of papers. However, authors from the largest connected component published the highest number of papers per component, 3.65 times more than the second highest number of papers per component. Articles written by authors from the largest component are published in 10 (out of 11) journals indexed by eLib, in all journals except "Visual mathematics". Additionally, the largest component is the only component that has a timespan which is the same as the lifetime of the whole network: it contains authors that published papers in 1932, as well as authors that published their first papers in 2011.

For each author in the co-authorship network can be characterized by several metrics used in social network analysis. Table 3 presents values of author metrics for top ten authors from the largest connected component when ranked by degree centrality. It can be seen that the best connected author is Ivan Gutman, a Serbian Academician from Kragujevac, who is connected to 50 other authors. Ivan Gutman also published the highest number of papers in the eLib journals. The best ranked author by betweenness centrality is Žarko Mijajlović, full professor at the Faculty of Mathematics, University of Belgrade, who is also the second best connected author. Mijajlović is the most central author in the largest connected component and can be viewed as the strongest middleman connecting different groups of authors. Top ten best ranked authors by degree centrality also have high betweenness centrality, i.e. all of them are positioned in the top 20 best ranked authors by betweenness centrality. Paul Erdős is the best ranked non-Serbian mathematician by betweenness centrality (his rank is 11). Petar M. Vasić has the highest value of clustering coefficient, which means that his co-authors established the tightest degree of collaboration between themselves compared to co-authors of other top ten highest degree authors. The highest value of author timespan (see Definition 2) has Stanković Bogoljub, a Serbian Academician from Novi Sad. He does not belong to the largest, but the second largest connected component. The first paper of Bogoljub Stanković published in the eLib journals is from 1953, while the last one is from 2011.

For each author in the eLib community we computed the Spearman correlation coefficient between the co-authorship based metrics and metrics of productivity. The number of published papers and author timespan are representatives of metrics of productivity, while degree centrality, betweenness centrality and clustering coefficient are co-authorship network based metrics. Results are summarized in Table 4. All computed correlations are significant at 0.05 level. It can be seen that there are strong correlations between the number of published papers and timespan, and between degree centrality and clustering coefficient. The correlation between the number of papers and timespan is expected to be strong: if an author has a large number of publications it is more likely that they are published in wider time range than in a smaller one. Strong correlation between degree centrality and clustering coefficient implies that co-authors of highly connected authors established a higher number of collaborations between themselves than co-authors of slightly connected authors. However, the most important are correlations between different types of author metrics, i.e. network-based metrics of importance (centrality metrics) and metrics of productivity (the number of published papers and timespan). It can be seen that betweenness centrality has stronger correlations with the number of published papers and timespan than degree centrality: correlations between betweenness and metrics of importance are moderate, while correlations between degree centrality and metrics of
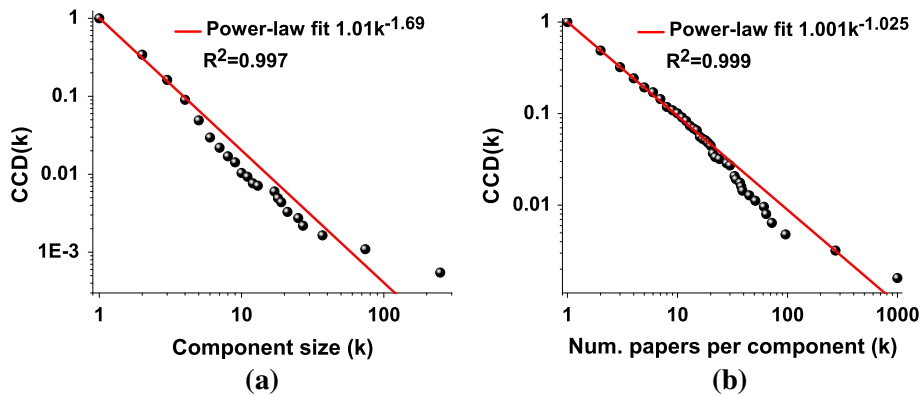
Fig. 6 The distribution of the size of components (**a**), and the number of papers per component (**b**)

Table 2 The ten largest connected components in the eLib co-authorship network

#N	#L	#P	#J	EP	$\langle d \rangle$	SW	D	CC
249	399	997	10	80	3.20	7.48	19	0.52
74	111	273	8	59	3.00	5.17	11	0.44
37	97	20	2	8	5.24	2.76	5	0.84
27	43	51	2	43	3.18	2.59	5	0.60
25	38	39	6	48	3.04	3.49	7	0.55
21	24	61	4	42	2.28	3.09	7	0.21
19	27	34	7	41	2.84	3.11	6	0.57
19	44	10	1	8	4.63	2.37	5	0.84
18	19	72	5	52	2.11	3.23	7	0.13
17	80	6	1	5	9.41	1.41	2	0.86

#N the number of nodes (authors), *#L* the number of links, *#P* the number of papers that authors in the component published, *#J* the number of journals where authors from the component published their papers, *EP* evolution period of the component, $\langle d \rangle$ average degree of node in component, *SW* small world coefficient, *D* diameter, *CC* clustering coefficient

importance are weak. This means that betweenness centrality is a better indicator of author productivity and long term presence in the eLib journals than degree centrality.

Links in the eLib co-authorship network have weights which denote the number of papers two authors jointly published. Therefore, link weight can be viewed as a measure of strength of collaboration between two authors. Figure 7a shows the distribution of link weights for the eLib co-authorship network plotted on semi-log scales. It can be seen that the distribution is monotonically decreasing. The majority of all links (64.69 %) have weight that is equal to one which implies that eLib authors mostly publish only one joint paper together in the eLib journals. The highest collaboration strength have authors Izidor Hafner and Tomislav Žitko. They published 23 joint papers in "Visual mathematics" in the period from 2002 to 2007. Another aspect related to scientific collaborations is timespan: the time passed from the first to the last publication of two authors. Figure 7a shows the distribution of link timespan for the eLib co-authorship network plotted on semi-log scales.

Table 3 The top ten highest degree authors in the largest component of the eLib co-authorship network

Name	Deg.	#P	#PR	S	SR	B	BR	CC
Ivan Gutman	50	71	1	40	10	12,263	7	0.0383
Žarko Mijajlović	16	44	4	40	10	17,983	1	0.1000
Dragoš Cvetković	14	41	5	42	9	12,255	8	0.0769
Zoran Ognjanović	14	23	14	16	31	4,089	15	0.1098
Boško Jovanović	12	36	6	37	12	6,581	10	0.0758
Slobodan Simić	12	21	15	36	13	3,042	17	0.1364
Miomir Stanković	11	12	23	34	15	3,218	16	0.1091
Petar M. Vasić	10	23	14	21	27	15,691	4	0.2000
Slaviša Prešić	8	26	11	47	5	16,722	2	0.1071
Jovan D. Kečkić	8	46	3	34	15	5,183	13	0.1428

Deg. degreee, *#P* the number of published papers, *#PR* rank of author according to the number of published papers, *S* author timespan, *SR* rank according to timespan, *B* betweenness, *BR* rank according to betweenness centrality, *CC* clustering coefficient

The largest link timespan is exhibited by authors Ranko Bojanić and Miloš Tomić. Those two authors published seven eLib papers together, the first in 1954, and the last in 1995.[7]

The importance of collaboration can be measured by link betweenness. Links with a high value of betweenness are separating different communities of nodes in a network, and this observation is used in the construction of the Girvan-Newman community detection algorithm (Girvan and Newman 2002). The highest value of betweenness in the eLib co-authorship network has the collaboration between Slaviša Prešić and Zoran Ivković. They published exactly one joint paper in 1967,[8] which can be considered as the most important paper for the overall connectedness of the largest connected component of the eLib co-authorship network.

For each link (collaboration) in the eLib community we computed the Spearman correlation coefficient between the following metrics: link weight, link timespan, number of authors in common for two authors that are connected by the link and link betweenness. Results are summarized in Table 5. All computed correlations, except the correlation between timespan and authors in common, are significant at 0.05 level. However, the only strong Spearman correlation is between link weight and link timespan. This means that authors who collaborated in a longer time interval tend to have more papers in common compared to co-authors with a shorter collaboration timespan. On the other hand, the absence of strong correlations between link weight and link betweenness indicates that non-frequent collaborations are equally important to the connectedness of components as frequent collaborations.

Evolution of the eLib co-authorship network

The eLib co-authorship network evolved from 11 isolated nodes (authors) in 1932–3597 nodes and 2,766 links in 2011. The first co-authorship link appeared in 1934 connecting

---

[7] The last joint eLib paper of Ranko Bojanić and Miloš Tomić is dedicated to the memory of Slobodan Aljančić with whom they co-authored their first eLib paper.

[8] The paper has title "Une simple méthode pour obtenir la décomposition effective de Wold dans le cas des chaines de Markoff de corélations stationnaires", and is published in "Matematički Vesnik".

Table 4 Values of Spearman's correlation coefficient for author metrics

	Num. papers	Timespan	Degree	Betweenness
Timespan	0.91			
Degree	0.12	0.11		
Betweenness	0.51	0.49	0.47	
Clustering coef.	−0.06	−0.07	0.79	0.11
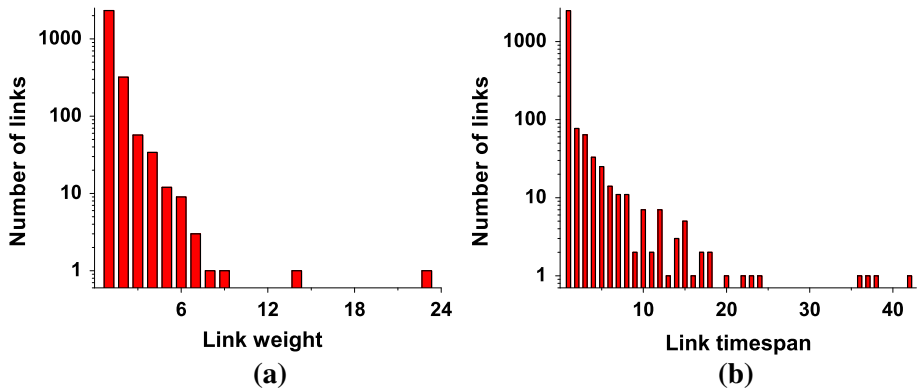
Fig. 7 The distribution of link weights (**a**) and link timespans (**b**)

authors Jovan Karamata and Hermann Wendelin who co-authored the paper titled "Zu Fragen über nichtvertauschbare Grenzprozesse." Figure 8 shows the evolution of the fraction of isolated nodes and the ratio between the number of links and non-isolated nodes. It can be observed that the fraction of isolated nodes is continuously decreasing after 1949:

- In 1949 92 % of the total number of authors have not collaborated with any other author by publishing papers in the eLib journals,
- In 1997 less than half of authors are isolated,
- In 2011 only 33 % of the total number of authors are those who exclusively publish single-authored papers in the eLib journals.

The evolution of the ratio between the number of links and non-isolated nodes (LNR) enables us to observe different periods in the evolution of the eLib co-authorship network that are characterized by different intensity of collaborations among eLib authors. If LNR increases as a network evolves then the number of links (collaborations) grows at a faster rate than the number of non-isolated nodes ("collaborative" authors), i.e. the degree of collaborative behaviour among authors is increasing. On the contrary, the decrease of LNR implies the decrease of collaborative behaviour. As can be seen in Fig. 8, six periods in the evolution of the eLib co-authorship network can be distinguished:

1. Before 1934, all published papers in the eLib journals are single-authored papers, which means that this period characterize the absence of collaborative behaviour among eLib authors.

Table 5 Values of Spearman's correlation coefficient for link (collaboration) metrics

|	Weight	Timespan	Authors in common
Timespan	0.78		
Authors in common	0.05	0.02	
Betweenness	0.17	0.19	−0.15
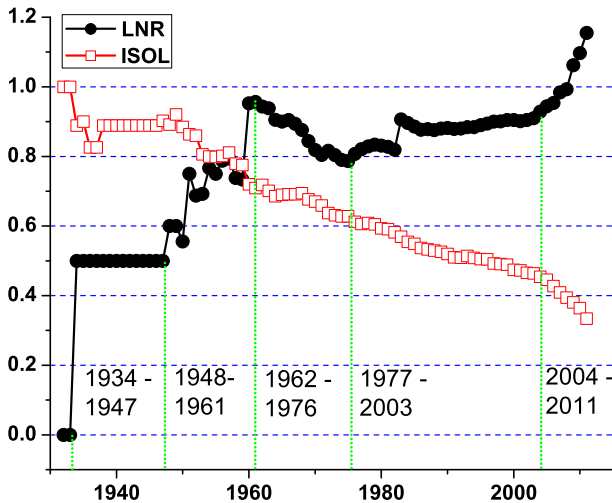
Fig. 8 The evolution of the ratio between the number of links and non-isolated nodes (LNR), and the fraction of isolated nodes (ISOL)

2. In the period from 1934 to 1947, collaborations among eLib authors started to emerge. In this period LNR is equal to 0.5 which means that all connected components have size 2. However, only two non single-authored papers were published in this period, that is four authors collaborated with others, and there were no authors who collaborated with more than one author.

3. The period from 1948 to 1961 was characterized by an intensive growth of collaborations among eLib authors: LNR increased from 0.5 to 0.96. The number of non-isolated authors by the end of 1961 is 66 and those authors are connected by 69 links, where 64 links (97 %) represent collaborations established in this period.

4. In the fourth period (1962–1976) LNR decreased from 0.95 to 0.81 implying lower intensity of collaborative behaviour among authors in comparison with the previous period. By the end of 1976 the number of non-isolated authors is 290. Those authors are connected by a significantly smaller number of links (234).

5. In the next period (1977–2003) LNR increased from 0.81 to 0.91 indicating a period when collaborations among authors again started to intensify. By the end of this period the network contained 1,151 non-isolated authors that were connected by 1,047 links.

6. The last period (2004–2011) has the same characteristics as the previous period (the number of links grows faster than the number of non-isolated nodes). However, in this period LNR grows at a faster rate than in all previous periods implying that the

Table 6 The number of collaborations between old authors (Old–Old), old and new authors (Old–New) and new authors (New–New) for the last four characteristic periods in the evolution of the eLib co-authorship network. The most dominant types of collaborations are bold

Period	Old–old	%	Old–new	%	New–new	%
1948–1961	39	**43.82**	19	21.35	31	34.83
1962–1976	66	32.04	79	**38.35**	61	29.61
1977–2003	372	**35.23**	331	31.34	353	33.43
2004–2011	487	23.08	543	25.73	1080	**51.18**

Bold indicates the most dominant type of collaboration in a specific period of time
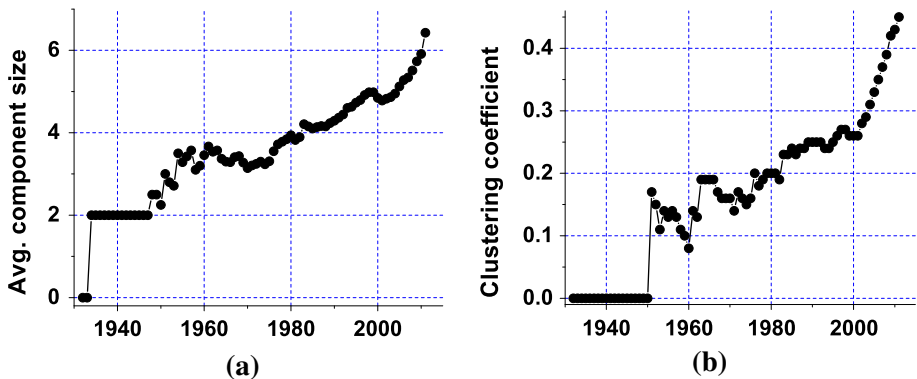

Fig. 9 The evolution of the average component size (**a**), and clustering coefficient (**b**) for non-trivial components

collaborative behaviour among eLib authors is the most intensive in the last years of eLib evolution. Additionally, in this period, for the first time in the evolution of the eLib co-authorship network, LNR became greater than one, denoting that the network contained more links than non-isolated nodes.

In order to determine the dominant type of collaboration for each of the last four characteristic periods in the evolution of the eLib co-authorship network, we computed the number of collaborations between old (returning) authors, old and new authors, and new authors. It is important to notice that in the computation of the number of collaborations link weights have to be considered (see Section 3). Table 6 shows the number of collaborations per type for different periods in eLib evolution. It can be seen that the dominant type of collaboration in periods 1948–1961 and 1977–2003 are collaborations between returning authors, in the period 1962–1976 collaborations between returning and new authors, and in the last years the majority of collaborations are formed by new authors. This means that the periods with different intensity of collaboration among eLib authors are additionally characterized by different types of collaborative behaviour.

As already mentioned, all components in the co-authorship network are either trivial or non-trivial. Since trivial components represent collaborations that resulted from publishing exactly one paper, their main characteristic is that they have not evolved in the examined time range (1932–2011). For each year in the network evolution we measured the average size (Fig. 9a) and clustering coefficient (Fig. 9b) of non-trivial components. It can be
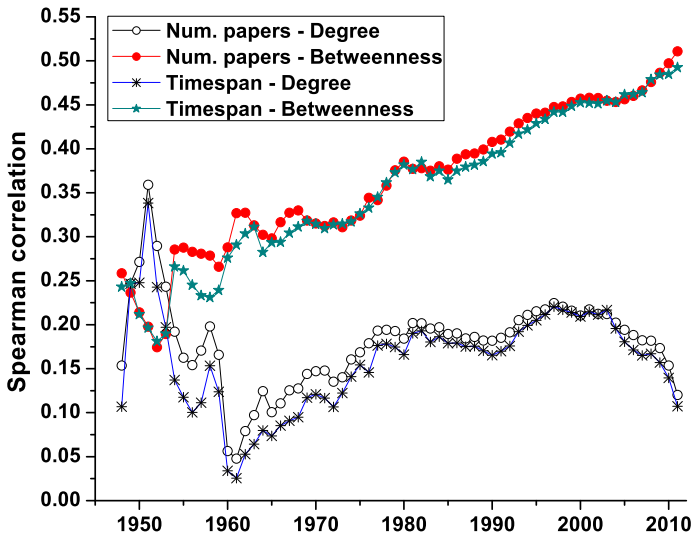
Fig. 10 The evolution of Spearman's correlations between co-authorship network based author metrics (degree and betweenness centrality) and metrics of productivity (the number of publications and author timespan)

observed that after 1970 connected components tend to be larger and more cohesive. The increase of average size and cohesiveness of non-trivial components corresponds to the increase of intensity of collaborations in the last two characteristic periods of the network evolution (see Fig. 8). The average size of non-trivial components in 1970 is equal to 3.14, while in 2011 it is more than two times larger (6.43). Similarly, the clustering coefficient of non-trivial components in 1970 is 0.16, while in 2011 it is nearly three times larger (0.45). Additionally, in the last decade of the network evolution (after 2001) both average size and clustering coefficient grow at higher rates than in previous years:

– The average size of non-trivial components increased from 3.14 in 1970 to 4.84 in 2000, which is average increase of 0.057 per year. In the last decade average increase per year of the average size of non-trivial components is 0.16 (2.8 times higher than in the period from 1970 to 2000).
– The clustering coefficient of non-trivial components increased from 0.16 in 1970 to 0.26 in 2000, which is average increase of 0.003 per year. The same quantity had average increase of 0.02 in the last decade of the network evolution.

As already shown in Section 4.2, betweenness centrality is a better indicator of author productivity and long-term presence in the eLib journals than degree in the co-authorship graph. We continue by presenting the analysis of the evolution of the strength of correlations between authors' centrality metrics and metrics of productivity and long-term presence in the eLib journals. For each year y in the evolution of the eLib co-authorship network, and each author A that was present in year y, we computed vector $V_y(A) = \langle n_y, t_y, d_y, b_y \rangle$, where n_y, t_y, d_y and b_y denote the number of papers, timespan, degree centrality and betweenness centrality of author A in year y, respectively. In that way a sequence of author metrics vectors per year is computed, which enabled us to investigate

the evolution of correlations between author metrics. During the whole examined period there is a strong correlation between author productivity and long-term presence: Spearman's correlation coefficient between the number of published papers and timespan is always greater than 0.9 after 1949. Figure 10 shows the evolution of Spearman's correlation coefficient between centrality metrics and metrics of productivity. It can be seen that only between 1950 and 1954 degree centrality had stronger correlations with productivity/long-term presence than betweenness centrality. More importantly, the strength of correlations between betweenness centrality and productivity/long-term presence is continuously increasing after 1970. Thus we can expect even stronger correlations between mentioned quantities in the future. To the contrary, the strength of correlations between degree and productivity/long-term presence started to decrease in 1997.

Another interesting aspect of the co-authorship network evolution is the change of the top-ranked author according to certain metric as the network evolves. Therefore, we determined the most productive, the best connected and the most central author for each five-year period in the evolution of the eLib co-authorship network. Results are summarized in Table 7. Mihailo Petrović (1868–1943), who is generally considered as one of the most prominent Serbian mathematician, is the most productive author in the first years of eLib evolution (1932–1945). In total he published 15 papers in the eLib journals, where three papers were published posthumously (after 1943). All of his eLib articles are single-authored papers. The next two decades are marked by the dominance of Jovan Karamata (1902–1967) in the eLib journals. From 1950 to 1970 this famous Serbian mathematician is at the same time the most productive, the most connected and the most central eLib author. He published 34 papers in the eLib journals in the period from 1932 to 1960. This means that his dominance lasted for a whole decade after he published his last paper. In the 1980s the most productive author is Djuro Kurepa (1907–1993), another famous Serbian mathematician known for his contributions to set theory and mathematical logic (especially the Kurepa tree). He published 45 papers in the eLib journals in the period from 1935 to 1989, but had only one collaborator among eLib authors (44 of his 45 papers published in the eLib journals are single-authored research works). In the first half of the first decade of the 21st century the most productive author is Bogoljub Stanković, a Serbian Academician born in 1924, who is still an active mathematician. He published 64 papers in the period from 1953 to 2011, and had 10 collaborators who are mathematicians mostly affiliated with the University of Novi Sad. From 1975 to 2000 the most, or one of the most, connected eLib authors is Petar M. Vasić who published 23 articles in the eLib journals in collaboration with 10 other authors. In 1985 the same author was the most central actor in the eLib co-authorship network. For other years in the period after 1975 the most central eLib authors are Slaviša Prešić (1933–2008) and his student Žarko Mijajlović. The most productive and the best connected author in the last years of eLib evolution is Ivan Gutman. According to the Mathematics Genealogy Project,[9] all eLib authors present in Table 7, except Djuro Kurepa who was a student of Maurice René Fréchet, are descendants of Mihailo Petrović.

The weight of a link in the eLib co-authorship network represents the number of papers two authors connected by the link published together in the eLib journals. Figure 11 shows the evolution of the average link weight for links contained in non-trivial connected components. It can be observed that only in two relatively short periods of time (from 1949 to 1956, and from 1976 to 1982) the average link weight exhibits an increasing trend. The increase of the average link weight denotes the intensification of already established

---

[9] http://www.genealogy.math.ndsu.nodak.edu/.

Table 7 The top ranked author according to the number of papers, degree, and betweenness centrality in different periods of eLib evolution

Year	Max. papers		Max. Degree		Max. betweenness	
1932	**3 authors**	2	–	0	–	0
1935	Mihailo Petrović	7	**2 authors**	1	–	0
1940	Mihailo Petrović	12	**4 authors**	1	–	0
1945	Mihailo Petrović	12	**4 authors**	1	–	0
1950	Jovan Karamata	16	Jovan Karamata	2	Jovan Karamata	1
1955	Jovan Karamata	28	Jovan Karamata	5	Jovan Karamata	25
1960	Jovan Karamata	34	Jovan Karamata	8	Jovan Karamata	61
1965	Jovan Karamata	34	Jovan Karamata	8	Jovan Karamata	61
1970	**2 authors**	34	Jovan Karamata	8	Jovan Karamata	68
1975	**2 authors**	34	Petar M. Vasić	9	Slaviša Prešić	113
1980	Djuro Kurepa	38	Petar M. Vasić	10	Slaviša Prešić	261
1985	Djuro Kurepa	42	**6 authors**	10	Petar M. Vasić	485
1990	Djuro Kurepa	45	**6 authors**	10	Slaviša Prešić	615
1995	Djuro Kurepa	45	**7 authors**	10	Slaviša Prešić	1,744
2000	Bogoljub Stanković	48	**7 authors**	10	Slaviša Prešić	2,482
2005	Bogoljub Stanković	55	Ivan Gutman	23	Slaviša Prešić	6,620
2010	Ivan Gutman	64	Ivan Gutman	45	Žarko Mijajlović	15,895
2011	Ivan Gutman	71	Ivan Gutman	50	Žarko Mijajlović	17,983

Bold emphasizes that there is more than one top-ranked author for a given year
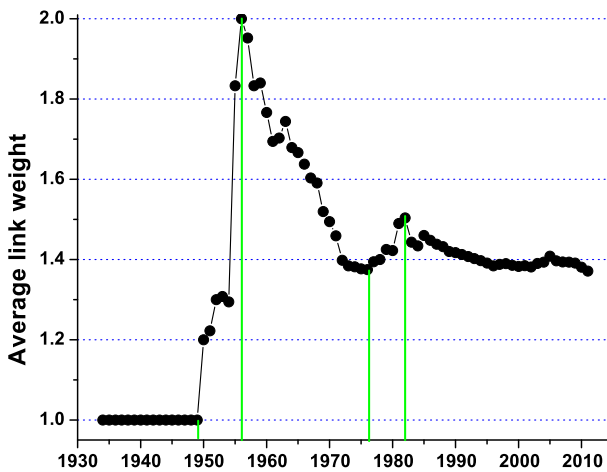
Fig. 11 The evolution of average link weight for non-trivial connected components

collaborations. For example, from 1949 to 1956 the average link weight increased from 1.0 to 2.0 meaning that each two authors who established their first collaboration before 1956 renewed the collaboration once again by the end of 1956 on average.

Conclusion

Scientific journals printed in Serbia are in the majority of cases covered insufficiently by the global and widely used digital libraries and bibliography database systems. The electronic library project of the Mathematical Institute of the Serbian Academy of Sciences and Arts (eLib) was started to fulfill this gap and provide online presence and long-term preservation of mathematical journals printed in Serbia. This electronic library contains articles published in 11 Serbian journals starting from 1932. Therefore, the bibliographic records contained in eLib enable a longitudinal scientometric study of a community which mainly consists of Serbian mathematicians.

In this paper we investigated the structure and evolution of the eLib co-authorship network that is determined by papers published in the eLib journals in the period from 1932 to 2011. Techniques and measures used in the analysis of social networks are employed in order to reveal structural properties and evolutionary trends in collaborations among eLib authors. Additionally, we provided the context in which the network formed by investigation of publication dynamics in the eLib journals (number of papers and authors per year), characteristics of published papers (number of authors per paper and fraction of single-authored papers) and characteristics of present authors (number of papers per author and fraction of returning authors).

Similarly to observations from previous research in co-authorship networks, the eLib co-authorship network is not a single connected graph. The analysis of connected components in the network revealed a topological diversity in the network structure that is characterized by the absence of a giant connected component, and power-law scaling behaviour regarding the size of components and the number of papers published by authors from the same component. Additionally, basic structural properties of co-authorship networks that are restricted to individual journals indexed by eLib are investigated in order to identify differences between them.

Evolutionary analysis of the eLib co-authorship network revealed that there are six different periods in the evolution of the network that are characterized by different intensity and type of collaborative behaviour among eLib authors. In the last two periods (from 1975) the intensity of collaborations exhibits a growing trend, and non-trivial connected components evolve in a way to become larger and more cohesive. This means that not only are new authors being integrated into non-trivial components, but also authors who have a co-author in common started to collaborate between themselves. Therefore, our findings for mathematical journals printed in Serbia are similar to those reported by Grossman (2002a) and Brunson et al. (2012) who observed a definite trend toward increasing collaboration in more recent times among mathematicians who publish their research work in journals indexed by *Mathematical Reviews*.

In this study we combined metrics used in analysis in social networks (degree and betweenness centrality, clustering and small-world coefficient) and metrics of productivity (the number of published papers and author timespan) to numerically represent characteristics of eLib authors. The analysis of author metrics showed that betweenness centrality is a better indicator of author productivity and long-term presence in the eLib journals than degree centrality. Additionally, evolutionary study of correlations between centrality and productivity metrics revealed that the strength of correlation between productivity metrics and betweenness centrality increases as the network evolves suggesting that even more stronger correlation can be expected in the future. We also investigated the change of the top-ranked eLib authors by co-authorship based metrics and metrics of productivity for each 5-year interval in the examined time period. Not surprisingly, eight widely recognized

Serbian mathematicians are identified as the top-ranked eLib authors in different periods of time, five of whom are/were members of the Serbian Academy of Sciences and Arts.

References

Amaral, L. A. N., Scala, A., Barthélémy, M., & Stanley, H. E. (2000). Classes of small-world networks. *Proceedings of the National Academy of Sciences of the USA*, *97*(21), 11,149–11,152.

Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509–512.

Barabási, A. L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A*, *311*, 590–614.

Batagelj, V., & Mrvar, A. (2000). Some analyses of Erdős collaboration graph. *Social Networks*, *22*(2), 173–186.

Bettencourt, L. M. A., Kaiser, D. I., & Kaur, J. (2009). Scientific discovery and topological transitions in collaboration networks. *Journal of Informetrics*, *3*(3), 210–221.

Börner, K., Maru, J. T., & Goldstone, R. L. (2004). The simultaneous evolution of author and paper networks. *Proceedings of the National Academy of Sciences of the USA*, *101*(Suppl 1), 5266–5273.

Brunson, J.C., Fassino, S., McInnes, A., Narayan, M., Richardson, B., Franck, C., Ion, P., & Laubenbacher, R. (2012). Evolutionary Events in a Mathematical Sciences Research Collaboration Network. arXiv: 1203.5158.

Cunningham, S., & Dillon, S. (1997). Authorship patterns in information systems. *Scientometrics*, *39*(1), 19–27.

de Solla Price, D. J. (1965). Networks of scientific papers. *Science*, *149*(3683), 510–515.

Ding, Y., Foo, S., & Chowdhury, G. (1998). A bibliometric analysis of collaboration in the field of information retrieval. *International Information and Library Review*, *30*(4), 367–376.

Dravec Braun, J. (2012). Effects of war on scientific production: Mathematics in Croatia from 1968 to 2008. *Scientometrics*, *93*(3), 931–936.

Garfield, E., & Sher, I. H. (1963). New factors in the evaluation of scientific literature through citation indexing. *American Documentation*, *14*(3), 195–201.

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, *99*(12), 7821–7826. doi:10.1073/pnas.122653799.

Gollapalli, S. D., Mitra, P., & Giles, C. L. (2011). Ranking authors in digital libraries. *In Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries* (pp. 251–254) New York, NY, USA: ACM, JCDL '11.

Grossman, J. (2002a). The evolution of the mathematical research collaboration graph. *Congressus Numerantium*, *158*, 201–212.

Grossman, J. (2002b). Patterns of collaboration in mathematical research. *SIAM News*, *35*(9), 8–9.

Grossman, J. W., & Ion, P. D. F. (1995). On a portion of the well known collaboration graph. *Congressus Numerantium*, *108*, 129–131.

Hou, H., Kretschmer, H., & Liu, Z. (2008). The structure of scientific collaboration networks in Scientometrics. *Scientometrics*, *75*(2), 189–202.

Ivanović, D., Surla, D., & Racković, M. (2012). Journal evaluation based on bibliometric indicators and the CERIF data model. *Computer Science and Information Systems*, *9*(2), 791–811.

Kalmukov, Y. (2012). Describing papers and reviewers' competences by taxonomy of keywords. *Computer Science and Information Systems*, *9*(2), 763–789.

Liben-Nowell, D., & Kleinberg, J. (2003). The link prediction problem for social networks. In *Proceedings of the Twelfth International Conference on Information and Knowledge Management* (pp. 556–559) New York, NY, USA: ACM, CIKM '03.

Lindsey, D. (1980). Production and citation measures in the sociology of science: The problem of multiple authorship. *Social Studies of Science*, *10*(2), 145–162.

Liu, X., Bollen, J., Nelson, M. L., & Van de Sompel, H. (2005). Co-authorship networks in the digital library research community. *Information Processing and Management*, *41*(6), 1462–1480.

Lotka, A. J. (1926). The frequency distribution of scientific production. *Journal of Washington Academy of Science*, *16*, 317–323.

Mijajlović, Ž., Ognjanović, Z., & Pejović, A. (2010). Digitization of mathematical editions in Serbia. *Mathematics in Computer Science*, *3*(3), 251–263.

Mimno, D., & McCallum, A. (2007). Mining a digital library for influential authors. In Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries (pp. 105–106) New York, NY, USA: ACM JCDL '07.

Nascimento, M. A., Sander, J., & Pound, J. (2003). Analysis of SIGMOD's co-authorship graph. *SIGMOD Record*, *32*(3), 8–10.

Newman, M. E. J. (2004a). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, *101*(1), 5200–5205.

Newman, M. E. J. (2001a). Scientific collaboration networks I: Network construction and fundamental results. *Physical Review E*, *64*(016), 131.

Newman, M. E. J. (2001b). Scientific collaboration networks II: Shortest paths, weighted networks, and centrality. *Physical Review E*, *64*(016), 132.

Newman, M. E. J. (2004b). Who is the best connected scientist? A study of scientific coauthorship networks. In E. Ben-Naim, H. Frauenfelder, & Z. Toroczkai (Eds.), *Complex networks* (Vol. 650, pp. 337–370). Lecture Notes in Physics Berlin Heidelberg: Springer.

Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, *46*(5), 323–351.

Pao, M. L. (1986). An empirical examination of Lotka's law. *Journal of the American Society for Information Science*, *37*(1), 26–33.

Perc, M. (2010). Growth and structure of Slovenia's scientific collaboration network. *Journal of Informetrics*, *4*(4), 475–482.

Rodriguez, M.A., Bollen, J. (2008). An algorithm to determine peer-reviewers. In Proceedings of the 17th ACM Conference on Information and Knowledge Management (pp. 319–328) New York, NY, USA: ACM, CIKM '08.

Voos, H. (1974). Lotka and information science. *Journal of the American Society for Information Science*, *25*(4), 270–272.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440–442.