

THE STRUCTURE AND PARAMETERS OF THE UNWEIGHTED CO-AUTHORSHIP NETWORK BASED ON DB REPEC DATA

S. V. Bredikhin, V. M. Lyapunov, N. G. Scherbakova

Institute of Computational Mathematics and Mathematical Geophysics SB RAS,
630090, Novosibirsk, Russia

DOI: 10.24412/2073-0667-2021-3-56-67

In this paper we investigate the structure of scientific collaboration based on co-authorship in publications retrieved from the RePEc database. The main attention is paid to the co-authorship network: nodes correspond to authors, and two authors are considered connected if they are coauthors of at least one publication. The network is represented by the undirected unweighted collaboration graph.

The co-authorship in articles provides a window on patterns of collaboration within the academic community. The study of scientific collaboration networks is carried out in two main directions. Empirical measurements provide detailed characteristics of the network: statistical properties of the distribution of degrees of nodes, global network parameters, nodes centralities (Newman 2001a; 2001b; 2001c). The study of the dynamic properties of real networks and network models makes it possible to identify the structural principles that govern the evolution of networks; dynamic properties, in turn, can explain static ones (Barabasi et al., 2001; Savić et al., 2017).

Co-authorship networks are used to identify research groups and the most important researchers and to predict their scientific success; to classify journals by degree of co-authorship and to make maps of science. Co-authorship networks can be constructed for different components of analysis, such as researchers, institutions, and countries. We use a researcher as the unit of analysis.

Let P be the set of publications under consideration and V — the set of the authors of these publications. We assume that each publication in P has at least two authors. For $i, j \in V$ let us define the binary relation of co-authorship R^{ca} : $i R^{ca} j \equiv (\exists p \in P)$ the author i is one of the authors of p and the author j is one of the authors of p . The co-authorship network is the weighted undirected network $N^{ca} = (V, E, w)$, V — the set of the authors, $E \subseteq V \times V$ — the set of weighted edges, $e = (i, j) \in E$, if $i R^{ca} j$. In the simplest case, $w(e) = 1$, regardless of how often these authors are coauthors and how much each contributed to coauthored publications, that is, the network can be considered unweighted. Then the graph corresponding to the unweighted network is represented by the $(0, 1)$ adjacency matrix $U = (u_{ij})$.

The articles used in this study were retrieved from the RePEc database (REPEC). In order to identify authors uniquely and to infer actual author identity we use the author „profile“ that authors create basing on the Author Service provided by the RePEc database (similar to Google Scholar). So we have $|P| = 91\,113$ publications with more than one author and $|V| = 32\,434$ authors of these publications. In total, these authors have published 364979 articles. The publications refer to the period from 1954 to 2019.

We presented the base statistics such as papers per author, authors per paper, maximal authors per paper and collaboration indices CI, DC, CC, MCC. We also measured the network parameters

This work was carried out under state contract with ICMMG SB RAS (0251-2021-0005).

such as the size of the largest component, network density, average distance, maximal distance and clustering coefficient. It was found that the distribution of numbers of coauthors for authors follows power law with exponent $\gamma = 1.3$.

The study showed that in the collection of publications under consideration the fraction of coauthored publications is small (25 %) and the prevailing trend is the presence of two coauthors in a publication (77 %). The most authors are indirectly connected to each other - the maximum component includes 90 % of authors. The co-authorship network under consideration is scale-free and shows the „small world“ effect.

Key words: bibliometrics, co-authorship, collaboration indices.

References

1. BEAVER D. D. Reflections on scientific collaboration (and its study): past, present and future // *Scientometrics*. 2001. V. 52. P. 365–377.
2. LAUDEL G. Who do we measure by co-authorship? // *Research Evaluation*. 2002. V. 11. P. 3–15.
3. STEPHAN P. How economics shapes science // Cambridge, Mass.: Harvard University Press, 2012. 384 pp.
4. DING Y. Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks // *J. of Informetrics*. 2011. V. 5. P. 187–203.
5. WALLACE M. L., LARIVIRE V., GINGRAS Y. A SmallWorld of citations? The influence of collaboration networks on citation practices // *PLoS ONE*. 2012. V. 7(3), e33339. [Electron. resource]. URL: <https://doi.org/10.1371/journal.pone.0033339>.
6. MARTIN T., BALL B., KARRER B., NEWMAN M. E. J. Coauthorship and citation patterns in the Physical Review // *Phys. Rev.* 2013. E 88, 012814. [Electron. resource]. URL: <https://doi.org/10.1103/PhysRevE.88.012814>.
7. NEWMAN M. E. J. The structure of scientific collaboration networks // *Proc. of the Nat. Acad. of Sci. of the USA*. 2001. V. 98(2). P. 404–409.
8. NEWMAN M. E. J. Scientific collaboration networks. I. Network construction and fundamental results // *Phys. Rev. E*, 64(1), 016131.
9. NEWMAN M. E. J. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality // *Phys. Rev. E*, 64(1), 016132.
10. BARABASI A. L., JEONG H., NEDA, Z. RAVASZ E., SCHUBERT A., V. CSEK T. Evolution of the social network of scientific collaborations // *arXiv: cond-mat/0104162* (2001).
11. Savić M., Ivanović M., Radovanović M., Ognjanović Z., Pejović A., Jakšić Kruger T. The structure and evolution of scientific collaboration in Serbian mathematical journals // *Scientometrics*. 2014. V. 101(3). P. 1805–1830. [Electron. resource]. URL: <https://doi.org/10.1007/s11192-014-1295-6>.
12. RePEc. General principles. [Electron. resource]. URL: <http://repec.org/>.
13. YOSHIKANE F., NOZAWA T., TSUJI K. Comparative analysis of co-authorship networks considering authors' roles in collaboration: Differences between the theoretical and application areas // *Scientometrics*. 2006. V. 68, iss. 3. P. 643–655.
14. BOUROS N., SOTIROPOULOS D.N., POURNARAKIS D., & GIAGLIS G.M. Social Network Analysis Within the ICMB Community: Co-Authorship Networks // *The 13th Intern. conf. on Mobile Business, ICMB 2014, London (UK), June 4–5, 2014*. P. 8. [Electron. resource]. URL: <http://aisel.aisnet.org/icmb2014/8>.
15. LAWANI S. M. Quality, collaboration and citations in cancer research: A bibliometric study // *Ph. D. Dissertation. Florida State Univ.*, 1980.
16. SUBRAMANYAM K. Bibliometric studies of research collaboration: A review // *J. Inform. Sci.* 1983. V. 6, iss. 1. P. 33–38.

17. AJIFERUKE I., BURELL Q., AND TAGUE J. Collaborative coefficient: A single measure of the degree of collaboration in research // *Scientometrics*. 1988. V. 14, iss. 5–6. P. 421–433.
18. EGGHE L. Theory of collaboration and collaborative measures // *Inform. Proc. and Manag.* 1991. V. 27, iss. 2–3. P. 177–202.
19. SARVANUR K. AND SRIKANTH R. Modified collaboration coefficient: A new measure for quantifying the degree of research collaboration // *Scientometrics*. 2010. V. 84, iss. 2. P. 365–376.
20. DISTEL P. *Teorya grafov*. Novosibirsk: Int-t matematiki, 2002. 336 s.
21. KHARARI F. *Teorya grafov*. M.: Mir. 1973. 297 s.
22. WATTS D.J. Networks, dynamics and the small-world phenomenon // *Amer. J. of Sociol.* 1999. V. 105, iss. 2. P. 493–527.
23. WATTS D.J., STROGATZ S.H. Collective dynamics of 'small-world' networks // *Nature*. 1998. V. 393. P. 440–442.
24. MILGRAM S. The small world problem // *Psychol. Today*. 1967. V. 2. P. 60–67.

СТРУКТУРА И ПАРАМЕТРЫ НЕВЗВЕШЕННОЙ СЕТИ СОВАВТОРСТВА НА ОСНОВЕ ДАННЫХ БД PERES

С. В. Бредихин, В. М. Ляпунов, Н. Г. Щербакова

Институт вычислительной математики и математической геофизики СО РАН,
630090, Новосибирск, Россия

УДК 001.12+303.2

DOI: 10.24412/2073-0667-2021-3-56-67

Исследуется структура сотрудничества на основе соавторства в научных публикациях. На основе информации, извлеченной из библиографической базы данных, построена сеть соавторства и измерены ее основные параметры. Показано, что сеть является масштабно-инвариантной и обладает структурой, свойственной „малому миру“. Вычислены индексы сотрудничества, отражающие вовлеченность ученых в совместную деятельность.

Ключевые слова: библиометрия, сеть соавторства, индексы сотрудничества.

Введение. Одним из классов комплексных сетей являются „сети сотрудничества“, в которых акторы связаны между собой благодаря членству в некоторой группе, например, спортсмены, занимающиеся одним видом спорта, актеры кино, участвующие в общем фильме; группы авторов, совместно подготовивших общую научную публикацию (НП) — соавторы.

Если более одного автора разделяет ответственность за НП, содержащую результаты конкретных исследований, то предполагается, что соавторы какое-то время работали сообща. Соавторство является достаточно хорошо документированным типом научного сотрудничества ученых. Общая ответственность за публикацию в солидном научном журнале редко возможна без какой-либо формы сотрудничества, которое, как правило, способствует эффективному решению поставленных в НП проблем, сокращая число заблуждений и ошибок.

Обзоры факторов, способствующих повышению роли, которую играет сотрудничество в научных исследованиях, а также побуждающих к нему мотиваций, можно найти в работах [1, 2], а также в книге [3]. Анализ сетей соавторства позволяет оценить степень взаимопроникновения идей, знаний и технологий и характеризует структуру научного сообщества. При этом могут рассматриваться различные уровни сотрудничества: между странами, в пределах одной страны, среди сотрудников организации, между научными направлениями, внутри одного направления, в рамках журнала или трудов конференции и т. д.

Сети цитирования и сети соавторства представляют разные взгляды на массив НП. Источником информации о НП, как правило, являются библиографические базы данных. При изучении сети цитирования НП являются ее узлами, а ориентированные связи между узлами строятся на основании цитирования. В сетях соавторства узлами являются авторы НП, а неориентированные связи строятся на основании того, что авторы являются

Исследования выполнены в рамках государственного задания ИВМ и МГ СО РАН (0251-2021-0005).

соавторами. Тем не менее, наблюдается взаимное влияние практик сотрудничества и цитирования ([4–6]). Например, результаты анализа параметров сети соавторства позволяют определить, насколько часто авторы цитируют своих соавторов по сравнению с другими цитируемыми авторами. Сети соавторства используются при выявлении групп ученых и наиболее важных направлений их деятельности и для прогнозирования успеха, позволяют классифицировать журналы по степени соавторства, составлять карты науки.

Изучение „сетей сотрудничества“ ведется по двум основным направлениям. Эмпирические измерения предоставляют детализированную характеристику сети: статистические свойства распределения степеней узлов, глобальные параметры сети, центральность [7–9]. Исследование динамических свойств реальных сетей и их моделей позволяет выявить структурные принципы, управляющие эволюцией сетей; динамические свойства, в свою очередь, могут объяснить статические [10, 11].

Данная работа содержит методы и результаты измерения статических параметров сети соавторства, представленной неориентированным невзвешенным графом, построенном на основе информации, извлеченной из библиографической базы данных (БД) RePEc [12]. Соавторство исследуется в рамках одной научной области.

1. Сеть публикаций и сеть соавторства. Обозначим $P = \{p_1, p_2, \dots, p_m\}$ множество НП, содержащихся в БД, $|P| = m$; $V = \{v_1, v_2, \dots, v_n\}$ — множество авторов НП, $|V| = n$.

Сеть публикаций (сеть авторства) $N^{\text{pub}} = (V, P, E^{\text{pub}})$ — это двудольная сеть с узлами, разделенными на два непересекающихся независимых (т.е. никакие два узла не являются смежными) множества V и P , при этом дуга $e = (v_i, p_j) \in E^{\text{pub}}$, если v_i является автором или соавтором НП p_j .

Матрица смежности (*матрица авторства*) $A = (a_{ij})$ графа, соответствующего N^{pub} , является матрицей размера $(n \times m)$, строки соответствуют авторам, а столбцы — публикациям:

$$a_{ij} = \begin{cases} 1, & \text{если } v_i \text{ является автором } p_j, \\ 0 & \text{в противном случае.} \end{cases} \quad (1)$$

Для $v_i, v_j \in V$ определим бинарное *отношение соавторства* R^{ca} :

$$v_i R^{\text{ca}} v_j \equiv (\exists p \in P), \text{ авторами } p \text{ являются } v_i, \text{ и } v_j. \quad (2)$$

Сетью соавторства будем называть взвешенную неориентированную сеть $N^{\text{ca}} = (V, E, w)$, в которой V — множество авторов, $E \subseteq V \times V$ — множество взвешенных ребер, $e = (v_i, v_j) \in E$, если выполняется $v_i R^{\text{ca}} v_j$. Узлами N^{ca} являются авторы НП, а в качестве ребер выступают пары авторов, являющихся соавторами по крайней мере одной НП. В простейшем случае $w(e) = 1$ независимо от того, насколько часто авторы v_i и v_j являются соавторами и какой вклад внес каждый автор, т.е. в этом случае сеть N^{ca} можно считать невзвешенной.

Тогда граф, соответствующий невзвешенной сети, представляется $(0, 1)$ -матрицей смежности $U = (u_{ij})$,

$$u_{ij} = \begin{cases} 1, & \text{если } v_i R^{\text{ca}} v_j, \\ 0 & \text{в противном случае.} \end{cases} \quad (3)$$

Матрицу U можно представить в виде произведения $A \cdot A^T$ с заменой всех ненулевых элементов на единицу. Здесь A — матрица смежности N^{pub} (1). Заметим, что сеть соавторства не всегда моделируется неориентированным графом. Например, если подчеркивается лидерство первого автора в списке соавторов, то сеть соавторства может быть представлена орграфом [13, 14].

Возможно, первым случаем использования сетевого подхода к понятию соавторства можно считать вычисление так называемых чисел Эрдеша (Erdős numbers) — расстояния в сети N^{ca} от какого-либо ученого, работающего в области математики, до венгерского математика Р. Erdős (1913–1996), автора около 1525 НП. Непосредственные соавторы Эрдеша, их 511, имеют число Эрдеша, равное единице.

Исследование невзвешенных N^{ca} -сетей показало, что типичное среднее расстояние между узлами приблизительно равно шести, что соответствует среднему расстоянию в сети социальных знакомств [9]. Сеть N^{ca} можно отнести к сетям „малого мира“, для которых среднее расстояние $L(N)$ растет логарифмически с ростом числа узлов, $L(N) \propto \log(n)$, а расстояние l между парой произвольно выбранных авторов растет пропорционально $\log(l)$.

2. Исходные данные. Проблема идентификации авторов на основе текста НП для БД значительного размера признана достаточно трудоемкой и зависящей от того, насколько хорошо документирована база данных. Так, например, трудности возникают при распознавании авторов НП, имеющих короткие обозначения ФИО, что приводит к необоснованным совпадениям и ошибочным выводам. Подход к идентификации авторов на основе анализа заголовка статьи, содержащего список авторов, места их работы и страны (с использованием библиотеки *Python difflib library*) можно найти в работе [6]. В нашем случае при отборе авторов и НП была использована информация, содержащаяся в БД. Во избежание путаницы, возникающей из-за различных форматов представления информации об авторе в НП, авторы создают „профили“ на основе сервиса *Author Service*, предоставляемого БД RePEc (подобно *Google Scholar*). В профиле автора указаны НП, размещаемые в БД, автором или соавтором которых он является. На момент 2020.01.31 число зарегистрировавших себя авторов составляет 70 549, среди них присутствуют и не задекларировавшие ни одной НП. Уникальных идентификаторов НП 470 333. Научных публикаций, задекларированных несколько раз — 138 882, эти НП предположительно имеют более одного автора. Дескрипторов НП, имеющих более одного автора и задекларированных тем же числом авторов, сколько указано в дескрипторе, 91 113. Число авторов этих НП 32 434. Всего этими авторами задекларировано 364 979 НП (в том числе и имеющих одного автора). Эти „фильтрованные“ данные взяты за основу при построении сети соавторства. Публикации относятся к периоду с 1954 по 2019 гг. Следует отметить, что дата создания БД RePEc относится к 1997 г.

Введем обозначения: S — множество авторов, имеющих публикации в соавторстве; Pa — множество публикаций, имеющих a авторов; P_+ — множество публикаций, имеющих более одного автора; P — множество всех публикаций, $P = P_1 \cup P_+$; $|P| = \sum_j |P_j|$; q — максимальное число соавторов публикаций из множества P .

Базовая статистическая информация о соавторстве в рассматриваемом множестве НП представлена в табл. 1. Таким образом, среднее число авторов НП множества P_+ равно 2,3, а именно приблизительно 77% всех НП множества P_+ имеют двух авторов, а примерно 20% — трех. На остальные группы приходится около 3% множества всех НП, выполненных в соавторстве. Значение данного параметра зависит, в частности, от рассматриваемой области науки (в данном случае — экономика). Пример для наиболее значимых размеров соавторства выглядит так (число соавторов, число статей): (2, 70 399); (3, 18 073); (4, 2245); (5, 251).

Среднее число НП в соавторстве, приходящихся на каждого автора, составляет 2,8. На рис. 1 приведено распределение продуктивности авторов множества S , опубликовавших

Таблица 1

Базовая информация о соавторстве

Параметр	Значение
Число авторов, имеющих НП в соавторстве	$ S = 32\,434$
Число НП, имеющих более одного автора	$ P_+ = 91\,113$
Число НП множества P , имеющих одного автора, для авторов из A	$ P_1 = 273\,866$
Максимальное число соавторов публикаций из множества P	$q = 17$
Последовательность значений числа соавторов публикаций	$\{2, 3, \dots, 12, 15, 16, 17\}$
Среднее число авторов НП из множества P_+	2,3
Среднее число НП из множества P_+ , приходящихся на одного автора	2,8

НП в соавторстве. По оси абсцисс указано число НП, по оси ординат — доля (в промилле от общего числа) авторов, опубликовавших указанное число НП, шкалы логарифмические. Распределение следует степенному закону $p_x \sim x^{-\gamma}$ с параметрами $x_{min} = 6$, $\gamma = 1,64$ (вычислено с помощью пакета `igraph`), т.е. большинство авторов имеет незначительное число НП, выполненных в соавторстве.

Отметим, что общее число НП растет со временем, в том числе растет и число НП, опубликованных в соавторстве, (период, число НП): (1954–1989, 3964); (1990–1999, 10 781); (2000–2009, 33 288); (2010–2019, 42941).

3. Эмпирические характеристики.

3.1. *Меры сотрудничества.* Для оценки степени сотрудничества и определения трендов в развитии сотрудничества разработан ряд мер. Рассмотрим некоторые из них, используемые для анализа соавторства в НП, относящихся к одной области науки.

Индекс сотрудничества (Collaborative Index [15]) определяется как

$$CI = \frac{\sum_{a=1}^q a \cdot |P_a|}{|P|}. \quad (4)$$

Это среднее число авторов публикаций из множества P . Заметим, что CI не равен нулю, если все НП имеют одного автора.

Степень сотрудничества (Degree of Collaboration [16]) определяется как

$$DC = 1 - \frac{|P_1|}{|P|}. \quad (5)$$

Это доля НП, имеющих более одного автора, по отношению ко всему множеству НП этих авторов. Множество значений лежит в интервале $[0, 1]$. Индекс равен нулю, если все НП имеют одного автора, и имеет тем большее значение, чем больше НП подготовлены в соавторстве.

Коэффициент сотрудничества (Collaborative Coefficient [17]) определяется как

$$CC = 1 - \frac{\sum_{a=1}^q |P_a|/a}{|P|}. \quad (6)$$

Таблица 2

Индексы сотрудничества для множества P

Мера сотрудничества	Значение
Индекс сотрудничества (4)	$CI = 1,315768$
Степень сотрудничества (5)	$DC = 0,249639$
Коэффициент сотрудничества (6)	$CC = 0,134959$
Модифицированный коэффициент сотрудничества (7)	$MCC = 0,134963$

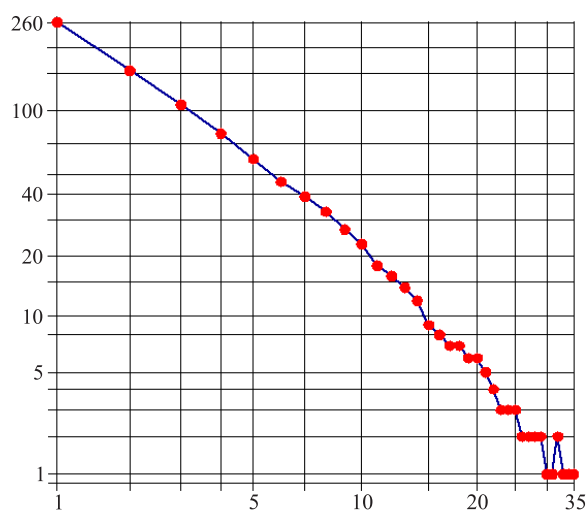
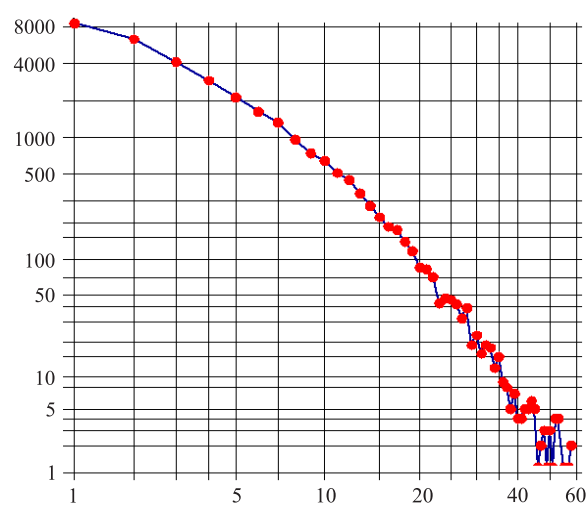
Рис. 1. Продуктивность авторов N^{ca} 

Рис. 2. Распределение степеней узлов сети

Коэффициент равен нулю, если в коллекции P нет публикаций в соавторстве. Дифференцирует уровни соавторства (2, 3 автора и т.д.). Заметим, что CC не равен единице при максимальном соавторстве, когда все авторы являются соавторами любой публикации коллекции.

Модифицированный коэффициент сотрудничества (Modified Collaboration Coefficient) [18, 19] определяется как

$$MCC = \frac{|S|}{|S| - 1} CC. \quad (7)$$

Это нормированный коэффициент сотрудничества. Предполагается, что $|S| > 1$. Коэффициент равен единице при максимальном числе соавторов и нулю, если все НП имеют одного автора.

В табл. 2 приведены значения рассмотренных индексов сотрудничества, вычисленных на основе анализа множества публикаций P . Согласно значениям индексов, уровень вовлеченности в совместную работу авторов публикаций множества P невысок.

3.2. Параметры сети соавторства. При построении сети соавторства мы рассматриваем множество авторов S и множество публикаций P_+ , каждая из которых имеет более одного автора. В невзвешенной сети соавторства $N^{ca} = (V, E^{ca})$ вершинам множества V соответствуют элементы множества S , ребро $e = (i, j) \in E^{ca}$, если выполняется (2) для $P \equiv P_+$. Размеры сети: $n = |V| = 32\,434$, $m = |E^{ca}| = 73\,969$. Матрица смежности $U = (u_{ij})$

Таблица 3

Параметры, определяющие структуру N^{ca}

Параметр	Значение
Распределение степеней узлов N^{ca} (число соавторов)	$p_x \sim x^{-\gamma}$, $x_{min} = 4$, $\gamma = 1,3$ (рис. 2)
Среднее значение степеней узлов	$\langle k \rangle = \langle \deg(\cdot) \rangle \sim 4,6$
Число связных компонент	1242
Размеры максимальной связной компоненты $N = (V, E)$	$ V = 29\,270$, $ E = 71\,780$, $\langle k \rangle \sim 4,9$
Вторая по числу узлов связная компонента	25 узлов
Реберная плотность в максимальной компоненте [20]:	$\Delta^U(N^{ca}) = 0,000168$
$\Delta^U(G) = \frac{2 E }{ V (V -1)}$	
Радиус [21]: $rad(G) = \min_{x \in V(G)} \max_{y \in V(G)} d(x, y)$	$rad(N^{ca}) = 13$
Диаметр [21]: $diam(G) = \max_{x \in V(G)} \max_{y \in V(G)} d(x, y)$	$diam(N^{ca}) = 25$
Характеристическая длина пути (среднее расстояние) [22]:	$L(N^{ca}) = 6,577475$
$L(G) = \frac{1}{n(n-1)} \sum_{x \in V} \sum_{y \neq x \in V} d(x, y)$	
В случайном графе $L_{rand} = \ln n / \ln \langle k \rangle$	$L_{rand} \sim 6,739148$
Локальный коэффициент кластеризации [23]:	$CC(N^{ca}) = 0,264428$
$CC(i) = \frac{\sum_{j,h} u_{ij}u_{ih}u_{jh}}{\deg(i)(\deg(i)-1)}$; $CC(N) = \frac{1}{n} \sum_i CC(i)$	
В случайном графе $C_{rand} = \langle k \rangle / n$	$C_{rand} \sim 0,000141$

сети N^{ca} организована в соответствии с равенством (3). В табл. 3. представлены основные параметры, характеризующие структуру N^{ca} .

На рис. 2 приведено распределение числа соавторов у авторов множества S . По оси абсцисс указано число соавторов, по оси ординат — число авторов, имеющих указанное число соавторов. Шкалы логарифмические. Распределение следует степенному закону $p_x \sim x^{-\gamma}$ с параметрами $x_{min} = 4$, $\gamma = 1,3$.

Число узлов в максимальной компоненте N^{ca} составляет примерно 90 % от всего множества узлов. Остальные компоненты имеют не более 25 узлов. Таким образом, большинство авторов к рассматриваемому моменту времени опосредованно связаны друг с другом. Реберная плотность достаточно низкая. Параметр $L(N)$ позиционируется в работе [22] как одна из характеристик сетей „малого мира“ [24], для которых характерно малое значение $L(N)$ (сравнимое со средним расстоянием в случайном графе) при достаточно большом значении n и небольшом среднем значении степеней вершин ($\langle k \rangle \ll n$). Такие сети также характеризуются более высокими, по сравнению со случайными графами, значениями локальных коэффициентов кластеризации. Сравнение значений параметров $L(N^{ca})$ и $CC(N^{ca})$ с соответствующими значениями параметров случайного графа позволяет заключить, что рассматриваемую сеть соавторства можно отнести к сетям „малого мира“.

Закключение. Анализ N^{ca} показал, что в рассматриваемой коллекции НП RePEc доля командной работы невелика и составляет приблизительно 25%. При этом преобладающей тенденцией является наличие двух соавторов (77% НП). Большинство авторов опосредованно связано друг с другом — максимальная компонента связности включает 90% авторов. Сеть соавторства является масштабно-инвариантной и относится к сетям „малого мира“. Дальнейшему исследованию подлежит взаимосвязь между соавторством и цитированием, а именно, получают ли НП, опубликованные в соавторстве, больше цитирований, чем опубликованные одним автором. Интерес представляют также параметры взвешенной сети соавторства, например, устойчивость авторских групп.

Список литературы

1. BEAVER D. D. Reflections on scientific collaboration (and its study): past, present and future // *Scientometrics*. 2001. V. 52. P. 365–377.
2. LAUDEL G. Who do we measure by co-authorship? // *Research Evaluation*. 2002. V. 11. P. 3–15.
3. STEPHAN P. How economics shapes science // Cambridge, Mass.: Harvard University Press, 2012. 384 pp.
4. DING Y. Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks // *J. of Informetrics*. 2011. V. 5. P. 187–203.
5. WALLACE M. L., LARIVIRE V., GINGRAS Y. A SmallWorld of citations? The influence of collaboration networks on citation practices // *PLoS ONE*. 2012. V. 7(3), e33339. [Electron. resource]. URL: <https://doi.org/10.1371/journal.pone.0033339>.
6. MARTIN T., BALL B., KARRER B., NEWMAN M. E. J. Coauthorship and citation patterns in the Physical Review // *Phys. Rev.* 2013. E 88, 012814. [Electron. resource]. URL: <https://doi.org/10.1103/PhysRevE.88.012814>.
7. NEWMAN M. E. J. The structure of scientific collaboration networks // *Proc. of the Nat. Acad. of Sci. of the USA*. 2001. V. 98(2). P. 404–409.
8. NEWMAN M. E. J. Scientific collaboration networks. I. Network construction and fundamental results // *Phys. Rev. E*, 64(1), 016131.
9. NEWMAN M. E. J. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality // *Phys. Rev. E*, 64(1), 016132.
10. BARABASI A. L., JEONG H., NEDA, Z. RAVASZ E., SCHUBERT A., V. CSEK T. Evolution of the social network of scientific collaborations // *arXiv: cond-mat/0104162* (2001).
11. Savić M., Ivanović M., Radovanović M., Ognjanović Z., Pejović A., Jakšić Kruger T. The structure and evolution of scientific collaboration in Serbian mathematical journals // *Scientometrics*. 2014. V. 101(3). P. 1805–1830. [Electron. resource]. URL: <https://doi.org/10.1007/s11192-014-1295-6>.
12. RePEc. General principles. [Electron. resource]. URL: <http://repec.org/>.
13. YOSHIKANE F., NOZAWA T., TSUJI K. Comparative analysis of co-authorship networks considering authors' roles in collaboration: Differences between the theoretical and application areas // *Scientometrics*. 2006. V. 68, iss. 3. P. 643–655.
14. BOUROS N., SOTIROPOULOS D.N., POURNARAKIS D., & GIAGLIS G.M. Social Network Analysis Within the ICMB Community: Co-Authorship Networks // *The 13th Intern. conf. on Mobile Business, ICMB 2014, London (UK), June 4–5, 2014*. P. 8. [Electron. resource]. URL: <http://aisel.aisnet.org/icmb2014/8>.
15. LAWANI S. M. Quality, collaboration and citations in cancer research: A bibliometric study // *Ph.D. Dissertation. Florida State Univ.*, 1980.

16. SUBRAMANYAM K. Bibliometric studies of research collaboration: A review // *J. Inform. Sci.* 1983. V. 6, iss. 1. P. 33–38.
17. AJIFERUKE I., BURELL Q., AND TAGUE J. Collaborative coefficient: A single measure of the degree of collaboration in research // *Scientometrics*. 1988. V. 14, iss. 5–6. P. 421–433.
18. EGGHE L. Theory of collaboration and collaborative measures // *Inform. Proc. and Manag.* 1991. V. 27, iss. 2–3. P. 177–202.
19. SARVANUR K. AND SRIKANTH R. Modified collaboration coefficient: A new measure for quantifying the degree of research collaboration // *Scientometrics*. 2010. V. 84, iss. 2. P. 365–376.
20. ДИСТЕЛЬ Р. Теория графов. Новосибирск: Ин-т математики, 2002. 336 с.
21. ХАРАРИ Ф. Теория графов. М.: Мир, 1973. 297 с.
22. WATTS D.J. Networks, dynamics and the small-world phenomenon // *Amer. J. of Sociol.* 1999. V. 105, iss. 2. P. 493–527.
23. WATTS D.J., STROGATZ S.H. Collective dynamics of 'small-world' networks // *Nature*. 1998. V. 393. P. 440–442.
24. MILGRAM S. The small world problem // *Psychol. Today*. 1967. V. 2. P. 60–67.



Бредихин Сергей Всеволодович — канд. техн. наук, зав. лабораторией Ин-та вычислительной математики и математической геофизики СО РАН; e-mail: bred@nsc.ru;

Сергей Бредихин окончил механико-математический факультет Новосибирского государственного университета в 1968 г. С 1968 г. — сотрудник Института автоматизации и электротехники СО РАН. Кандидат технических наук с 1983 г. С 1988 г. — заведующий Лабораторией прикладных систем Института вычислительной математики и математической геофизики СО РАН. Являлся техническим директором проекта „Сеть Интернет Новосибирского Научного Центра“. Лауреат государственной премии по науке и технике 2012 г. В сфере его научных интересов — измерение и анализ сетей распределенных информационных структур. Автор и соавтор более 110 работ и двух монографий: „Методы библиометрии и рынок электронной научной периодики“, „Анализ цитирования в библиометрии“.

Sergey Bredikhin graduated from Novosibirsk State University in 1968 (faculty of Mechanics and Mathematics). In 1968 he became an employee of Institute of Automation and Electrometry SB RAS. In 1983 he received PhD degree in Engineering Science. Since 1988 he is the head of Applied Systems laboratory of Institute of Computational Mathematics and Mathematical Geophysics SB RAS. He was the technical manager

of „Akademgorodok Internet Project“. He is the state prize winner in science and engineering (2012). Sphere of his scientific interests — the measurement and analysis of networks of the distributed information structures. He is the author and co-author of more than 110 works and two monographs: „Metody bibliometrii i rynek elektronnoy nauchnoy periodiki“, „Ansliz tsitirovaniya v bibliometrii“.



Ляпунов Виктор Михайлович — ведущий инженер Ин-та вычислительной математики и математической геофизики СО РАН; e-mail: vic@nsc.ru;

Виктор Ляпунов окончил механико-математический факультет Новосибирского государственного университета в 1978 году. В 1978 г. стал сотрудником Вычислительного Центра СО АН СССР, а с 1990 г. — сотрудником Института систем информатики СО АН СССР. С 2004 г. — ведущий инженер Института вычислительной математики и математической геофизики СО РАН. Занимается вопросами извлечения информации из баз данных и обработкой больших массивов данных. Соавтор более 10 работ в этой области.

Victor Lyapunov graduated from Novosibirsk State University in 1978 (faculty of Mechanics and Mathematics). In 1978, he became an employee of Computing Center of SB AS USSR, since 1990 — an employee of Institute

of Informatics Systems SB RAS. Since 2004 he works as software engineer in Institute of Computational Mathematics and Mathematical Geophysics SB RAS. His current research interests include methods of information extracting from databases and processing of large data sets. He is the co-author of more than 10 works in that area.



Щербакова Наталья Григорьевна — ст. науч. сотр. Ин-та вычислительной математики и математической геофизики СО РАН; e-mail: nata@nsc.ru.

Наталья Щербакова окончила Новосибирский государственный университет по специальности „Математическая лингвистика“ в 1967 г. С 1967 г. работала в Институте математики СО РАН, затем в Институте автоматизации и электрометрии СО РАН в области создания программного обеспечения систем передачи данных. С 2000 г. — сотрудник Института вычислительной математики и математической геофизики СО РАН, где с 2002 г. занимает должность старшего научного сотрудника. Являлась участником проекта „Сеть Интернет

Новосибирского Научного Центра“, занималась вопросами мониторинга и анализа IP-сетей. Автор и соавтор более 40 работ, соавтор монографии „Анализ цитирования в библиометрии“. Текущие интересы лежат в области исследования методов оценки научной деятельности на основе анализа цитирования научной литературы.

Natalia Shcherbakova graduated from Novosibirsk State University in 1967 (mathematical linguistics). Since 1967 she worked at Institute of Mathematics SB RAS, then at Institute of Automation and Electrometry SB RAS in the field of software design for data transmission systems. In 2000 — the employee of Institute of Computational Mathematics and Mathematical Geophysics SB RAS, since 2002 works as senior researcher. She is a member of „Akademgorodok Internet Project“, dealt with software of monitoring and the analysis of IP networks. She is the author and co-author of more than 40 works, the co-author of the monograph „Ansliz tsitirovaniya v bibliometrii“. The current research interests lie in the field of bibliometrics: methods of measuring of scientific activity on the base of citations.

Дата поступления — 16.03.2021