

RANKING AUTHORS OF THE WEIGHTED CO-AUTHORSHIP NETWORK: ANALYSIS OF 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: 10.24412/2073-0667-2021-4-67-83

In the previous paper [12] we investigated the co-authorship network (N^{ca}) represented by an unweighted graph: nodes correspond to authors, and two authors are considered connected if they are coauthors of at least one publication. Basic network properties are: existence of the giant component (includes 90 % of authors), “small worldness” [24] and a power-law fitting of the distribution of coauthors. In this paper we focus on centrality measures in order to identify key authors on the base of the weighted co-authorship network. Using co-authorship data from the distributed database RePEc [13] we construct two weighted networks that differ in the way of computing edge weights.

Let P ($|P| = l$) be the set of publications and assume that each publication in P has at least two authors. Let V ($|V| = n$) be the set of authors of these publications and $a_{ij} = 1$ if i is the author of the publication j . For the network N_T^{ca} the strength of the collaborative tie (the edge weight)

between the authors i and j is equal to the number of joint papers (T -method): $w(i, j) = \sum_{k=1}^l a_{ik} \cdot a_{jk}$. For the network N_F^{ca} the edge weight between the authors i and j depends not only on the number of coauthored papers, but also on the number of other coauthors of these papers (F -method [7]):

$w(i, j) = \sum_{k=1}^l \frac{a_{ik} \cdot a_{jk}}{n_k - 1}$, where n_k is the number of authors of the publication k .

The raw data processing procedure is presented in [12], as a result the number of authors $|V| = 32\,434$ and the number of coauthored publications $|P| = 91\,113$. For each of the network N^{ca} , N_T^{ca} , N_F^{ca} four measures of centrality such as degree, closeness, betweenness and eigenvector have been calculated and the tables (tabs. 2–4) containing the names of the authors with the highest ranks are provided. It should be noted that these authors have high h-index values (according to Google Scholar search engine or IDEAS ranking system [25] based on all publications of the authors).

In order to study the dependence of author ranks on the method of calculating the contributions of authors to publications we calculated Pearson’s correlation coefficients and Spearman’s rank correlation coefficients for the same centrality measures for the networks under consideration. It was shown that regardless of how the edge weights are calculated the same centrality measures have significant correlation with each other. The most significant correlation according to both coefficients is fixed for the betweenness centrality, the least — for the eigenvector centrality, which determines the “prestige” of the network actor.

To illustrate the studied ways of calculating edge weights and the dependence of node ranks on the method and a node location, we considered the 12-node component of N^{ca} and applied four centrality measures to its weighted representations. We see that the ranks of authors differ depending on the method of edge weights calculating. On the base of node ranks we calculated node weights

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

and presented new ranks of authors (tab. 10) within any component representation and centrality measure used. It is noted that the high ranked authors are the influential persons with a large number of citations.

The purpose of further research is to identify the relationship between key authors and the number of citations of coauthored publications. The question of interest is whether collaborative publications receive more citations than single author publications.

Key words: bibliometry, co-authorship network, centrality measures, key authors.

References

1. EVERETT M. G., BORGATTI S. P. The centrality of groups and classes // *J. of Math. Sociology*. 1999. V. 23, iss. 3. P. 181–201.
2. BORGATTI S. P. Identifying sets of key players in a social network // *Comput. Math. Organiz. Theory*. 2006. V. 12. P. 21–34.
3. BOLLEN J., RODRIGUEZ M. A., VAN DE SOMPEL H. Journal status // *Scientometrics*. 2006. V. 69, iss. 3. P. 669–687.
4. LEYDESDORFF L. Betweenness centrality as an indicator of the interdisciplinarity of scientific journals // *J. of the Amer. Soc. for Inform. Sci. and Technol.* 2007. V. 58, iss. 9. P. 1303–1319.
5. PENG T-Q. Assortative mixing, preferential attachment, and triadic closure: A longitudinal study of tie-generative mechanisms in journal citation networks // *J. of Informetrics*. 2015. V. 9, iss. 2. P. 250–262.
6. NEWMAN M. E. J. Scientific collaboration networks. I. Network construction and fundamental results // *Phys. Rev. E*. 2001. V. 64, iss. 1. 016131.
7. NEWMAN M. E. J. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality // *Phys. Rev. E*. 2001. V. 64, iss. 1. 016132.
8. NEWMAN M. E. J. Who is the best connected scientist? A study of scientific coauthorship networks // *Complex network. Lect. notes in Phys.* 2004. V. 650. P. 337–390.
9. YAN E., DING Y. Applying centrality measures to impact analysis // *J. of Amer. Soc. for Inform. Sci. and Technol.* 2009. V. 60, iss. 10. P. 2107–2118.
10. UDDIN S., HOSSAIN L., ABBASI A., RASMUSSEN K. Trend and efficiency analysis of co-authorship network // *Scientometrics*. 2012. V. 90. P. 687–699.
11. YOUNGBLOOD M., LAHTI D. A bibliometric analysis of the interdisciplinary field of cultural evolution // *Palgrave Communications*. 2018. Art. 120.
12. BREDIKHIN S. V., LYAPUNOV V. M., SCHERBAKOVA N. G. The structure and parameters of the unweighted co-authorship network based on DB RePEc data // *Problems of informatics*, 2021 (in print)
13. REPEC. General principles: [Electron. Res.]: <http://repec.org/>.
14. PERIANES-RODRIGUEZ A., WALTMAN L., VAN ECK N. J. Constructing bibliometric networks: A comparison between full and fractional counting // *J. of Informetrics*. 2016. V. 10, iss. 4. P. 1178–1195.
15. BREDIKHIN S.V., LYAPUNOV V. M., SHCHERBAKOVA N. G. Bibliometricheskie seti nauchnykh statey i zhurnalov. Novosibirsk, ICM&MG SB RAS, 2021, 334 p. [Electron. Res.]: <https://www.elibrary.ru/item.asp?id=45606936>.
16. NIEMINEN J. On centrality in a graph // *Scandinav. J. of Psych.* 1974. V. 15. P. 322–336.
17. BARRAT A., BARTHELEMY M., PASTOR-SATORRAS R., VESPIGNANI A. The architecture of complex weighted networks // *Proc. of the Nation. Acad. of Sci.* 2004. V. 101, iss. 11. P. 3747–3752.
18. BAVELAS A. Communication patterns in task-oriented groups // *J. of the Acoustical Soc. of Amer.* 1950. V. 22. P. 271–288.
19. BEAUCHAMP M. A. An improved index of centrality // *Behav. Sci.* 1965. V. 10. P. 161–163.

20. ANTHONISSE J. M. The rush in a directed graph // Technic. rep. BN 9/71. Amsterdam: Stichting Mathematisch Centrum, 1971.
21. FREEMAN L. C. A set of measures of centrality based upon betweenness // Sociometry. 1977. V. 40. P. 35–41.
22. BONACICH P. Factoring and weighting approaches to status scores and clique identification // J. of Math. Sociol. 1972. V. 2. P. 113–120.
23. WATTS D. J. Networks, dynamics and the small-world phenomenon // Amer. J. of Sociol. 1999. V. 105, iss. 2. P. 493–527.
24. WATTS D. J., STROGATZ S. H. Collective dynamics of 'small-world' networks // Nature. 1998. V. 393. P. 440–442.
25. IDEAS [Electron. Res.]: <https://ideas.repec.org/top/top.person.hindex.html>.

РАНЖИРОВАНИЕ УЗЛОВ ВЗВЕШЕННОЙ СЕТИ СООВТОРСТВА: АНАЛИЗ ДАННЫХ БД RERES

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

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

УДК 519.177

DOI: 10.24412/2073-0667-2021-4-67-83

Изучаются взвешенные сети соавторства, построенные на основе извлеченной из библиографической БД информации. Узлами сетей являются авторы научных публикаций, а связи устанавливаются на основе бинарного отношения соавторства. Рассматриваются два метода установления весов ребер, отражающие вклад каждого соавтора, если эти сведения не указаны явно. Вычислены меры центральности по степени, близости, посредничеству и собственному вектору. Приведены результаты ранжирования авторов, исследована их зависимость от выбора метода назначения весов. Показано, что меры центральности имеют высокий уровень корреляции, наивысший для центральности по посредничеству.

Ключевые слова: библиометрия, сеть соавторства, методы определения вкладов авторов, меры центральности.

Введение. Одним из направлений изучения научного сотрудничества является анализ сетей соавторства, которые конструируются на основе информации, извлекаемой из библиографических БД. Важным является тот факт, что акт научного сотрудничества явно задокументирован в совместных публикациях авторов. Узлами сети соавторства (N^{ca}) являются научные публикации (НП), а в качестве ребер выступают пары авторов НП (соавторы), которые считаются связанными, если оба являются авторами хотя бы одной совместной НП. Сетевой подход позволяет зафиксировать растущую структуру научного сообщества в определенный момент времени и оценить вклад отдельных ученых в ее создание.

Проблема выявления ключевых игроков в социальных сетях в значительной мере была изучена в работах Ф. Боначича и Л. Фримана периода 1972–1979 гг., которые ввели понятие “центральности” узлов и разработали методы, позволяющие ее вычислять. Таким образом появилась возможность оценивать структурную важность участников сети. В работе [1] определения “стандартных” мер сетевой центральности отдельных авторов (степень, близость, посредничество) расширены на группы авторов, что позволяет отвечать на вопрос о том, насколько центральное место в сети занимает та или иная группа. Развитие предложенных понятий позволило оценивать центральность (или периферийность) групп узлов и их важность для сети в целом [2]. Статус научного журнала (НЖ), как правило, определяется двумя факторами: общим числом полученных цитирований (популярностью), которые получает НЖ, и престижем цитирующих НЖ. Авторы статьи [3]

Исследования выполнены в рамках государственного задания ИВМиМГ СО РАН (0251-2021-0005).

развивают понятие статуса для получения экспертной оценки НЖ. Работа [4] демонстрирует результаты применения ряда мер центральности узлов для анализа междисциплинарных направлений и составления карт науки. В статье [5] аппарат мер центральности применяется для исследования организационных принципов, лежащих в основе развития сетей цитирования НЖ.

Изучение статистических свойств N^{ca} -сетей, построенных на основе информации, извлеченной из различных библиографических БД, начато в серии работ М. Ньюмана [6–8]. Опираясь на результаты измерения мер центральности, автор предлагает ряд методов, позволяющих аргументированно отвечать на вопрос о том, кто из ученых в своей области исследований обладает наиболее широкими связями. В статье [9] рассматривается процесс роста конкретной сети N^{ca} с целью изучения шаблонов соавторства. Основываясь на результатах вычисления мер центральности, авторы приходят к заключению о высокой степени корреляции полученных значений с числом цитирований. Взаимосвязь мер центральности с эффективностью НП изучается также в коллективной работе [10]. Авторы работы [11] сочетают методы библиометрического и кластерного анализов для изучения междисциплинарных областей культурной эволюции. Для исследования тенденций сотрудничества и коллективной продуктивности используются две меры центральности — по близости и собственному вектору — как в большей степени определяющие структурные изменения в сети. Результаты исследований подтверждают появление разнообразных междисциплинарных областей.

В настоящей работе, являющейся продолжением работы [12], исследуется влияние способов определения вкладов авторов на центральность узлов сети с использованием четырех мер центральности: по степени, близости, посредничеству и собственному вектору. Вычислительный эксперимент основан на библиографической информации о НП, проиндексированных в БД RePEc [13].

1. Методы анализа. 1.1. *Измерение вкладов авторов* Веса ребер сети соавторства базируются на весовых схемах, отражающих вклады авторов в НП. В данном случае применяются два варианта определения вкладов авторов: *полный (стандартный)* и *частичный*, которые обозначим *T-метод* и *F-метод* соответственно. При использовании *T-метода* предполагается, что вес НП, выполненной в соавторстве, равен сумме вкладов всех соавторов из расчета по единице для каждого соавтора, т. е. вес НП равен числу соавторов. В свою очередь, при использовании *F-метода* предполагается, что вес выполненной в соавторстве НП k равен сумме дробей $1/n_k$, где n_k — число соавторов НП k . Таким образом, при расчете веса научной публикации *F-методом* все НП имеют одинаковый вес, равный единице.

Пусть V ($|V| = n$) — множество авторов, а P ($|P| = l$) — множество публикаций, авторами (соавторами) которых являются авторы из V . И пусть $A = (a_{ij})$ — матрица авторства размера $n \times l$, $a_{ij} = 1$, если i является автором НП j . Матрицу авторства можно рассматривать как матрицу смежности ориентированной двудольной сети публикаций (авторства) $N^{\text{pub}} = (V, P, E)$, связывающей авторов с их публикациями. Пусть $U = (u_{ij})$ — матрица соавторства размера $n \times n$, элемент u_{ij} равен числу соавторских связей между учеными i и j , т. е. числу НП, авторами которых являются i и j . Тогда

$$u_{ij} = \sum_{k=1}^l a_{ik} \cdot a_{jk}. \quad (1)$$

В матричной нотации: $U = A \cdot A^\top$.

Таким образом, матрицу соавторства можно получить путем умножения матрицы авторства на транспонированную матрицу авторства. Диагональный элемент u_{ii} матрицы U равен числу НП, в которых участвовал автор i . Ввиду того что единственный автор НП не считается своим соавтором, диагональные элементы матрицы U устанавливаются равными нулю. Матрицу U можно рассматривать как матрицу весов $W = (w_{ij})$, $w_{ij} = u_{ij}$ сети соавторства $N^{ca} = (V, E, W)$, связывающей авторов на основе соавторства в НП. Вычисление весов ребер по T -методу осуществляется согласно (1). При использовании этого метода можно применять нормирование, например, согласно схеме Солтона, вес ребра (i, j) равен $w_{ij} = \frac{u_{ij}}{\sqrt{|P_i| \cdot |P_j|}}$, где P_i, P_j — множество публикаций, авторами которых являются i, j соответственно.

Перейдем к F -методу, который был предложен в работе [7]. Матрицу соавторства обозначим $U^* = (u_{ij}^*)$. Значение u_{ij}^* , указывающее на соавторство между учеными i и j , определяется так:

$$u_{ij}^* = \sum_{k=1}^l \frac{a_{ik} \cdot a_{jk}}{n_k - 1}, \quad (2)$$

где n_k — число авторов НП k . Отметим, что в знаменателе (2) из n_k вычитается единица, поскольку автор не считается своим соавтором ([14]). В матричной нотации: $U^* = A \cdot \text{diag}(A^\top \mathbf{1} - \mathbf{1})^{-1} \cdot A^\top$, где $\text{diag}(v)$ — диагональная матрица с элементами вектора \mathbf{v} , размещенными на главной диагонали, $\mathbf{1}$ — вектор-столбец, все n элементов которого равны 1. Элементы на главной диагонали матрицы соавторства U^* устанавливаются равными нулю. Матрицу U^* можно рассматривать как матрицу весов $W = (w_{ij})$, $w_{ij} = u_{ij}^*$, сети соавторства $N^{ca} = (V, E, W)$. Вычисление весов ребер по F -методу осуществляется согласно (2).

1.2. *Меры центральности.* Для определения авторов, играющих ключевую роль в сети, используем четыре меры центральности вершин графа, представляющего сеть (см. гл. 4 монографии [15]). Рассмотрим неориентированный связный граф $G = (V, E)$, $|V| = n$, без кратных ребер и петель с матрицей смежности $A = (a_{ij})$.

1.2.1. Центральность по степени [16] определяется числом связей (ребер) узла с его непосредственными соседями. Центральность $C_D(i)$ узла i определяется равенством

$$C_D(i) = \sum_{j=1}^n a_{ij}. \quad (3)$$

Как правило, узлы с большим числом инцидентных ребер являются наиболее важными и имеют потенциальную возможность влиять на ближайших соседей и опосредованно на остальные узлы. Заметим, что в качестве такого узла сети N^{ca} может оказаться автор НП, имеющей значительное число соавторов лишь одной НП.

Взвешенная центральность по степени $C_D^w(i)$ узла i , согласно [17], определяется равенством

$$C_D^w(i) = \sum_{j=1}^n a_{ij} \cdot w_{ij}. \quad (4)$$

1.2.2. Центральность по близости $C_C(i)$ узла i определяется на основе суммы геодезических расстояний от i до остальных узлов сети (или от всех узлов до узла i). Эта мера

указывает, как i влияет на остальные узлы, используя кратчайшие пути. В терминах передачи информации $C_C(i)$ можно интерпретировать как скорость ее распространения. Согласно [18], мера $C_C(i)$ определяется как число, обратное к сумме расстояний:

$$C_C(i) = \frac{1}{\sum_j d_{ji}}, \quad (5)$$

где d_{ji} — расстояние от узла j до узла i .

Относительная центральность по близости, согласно [19], определяется равенством

$$C'_C(i) = \frac{n-1}{\sum_j d_{ji}}. \quad (6)$$

Для взвешенного графа равенство (5) формулируется так:

$$C_C^w(i) = \frac{1}{\sum_j d_{ji}^w}, \quad (7)$$

где d_{ji}^w — взвешенное расстояние.

1.2.3. Центральность по посредничеству $C_B(i)$ базируется на числе кратчайших путей, проходящих через узел i . Узел сети с большим значением параметра может рассматриваться как мост, связывающий различные группы узлов. В терминах передачи информации параметр можно рассматривать как узел, контролирующий ее распространение. Согласно [20, 21], мера $C_B(i)$ определяется равенством

$$C_B(i) = \sum_{j=1}^n \sum_{k=1}^n \frac{g_{jk}(i)}{g_{jk}} \quad (i \neq j \neq k). \quad (8)$$

Здесь g_{jk} — число всех кратчайших путей между узлами j и k , $g_{jk}(i)$ — число таких путей, проходящих через узел i . Для взвешенных сетей при выборе кратчайших путей учитывается семантика, вложенная в определение веса ребра. Так, если предпочтительными являются большие значения весов ребер (отражающих качество связей), то при вычислении длины пути вместо весов ребер, как правило, используются обратные величины (w_{ij} заменяется на $1/w_{ij}$).

1.2.4. Центральность по собственному вектору $C_E(i)$ определяет “престиж” узла i в сети. Эта мера присваивает относительные оценки всем узлам в сети на основе того принципа, что соединения с узлами с высокими оценками вносят больший вклад в оценку рассматриваемого узла, чем равные соединения с узлами с низкими показателями. В частности, центральность собственного вектора узла пропорциональна сумме центральностей собственных векторов его соседей:

$$C_E(i) = \frac{1}{\lambda} \sum_{j=1}^n a_{ji} \cdot C_E(j), \quad (9)$$

где λ — константа.

На практике значения $C_E(i)$ определяются путем вычисления собственного вектора, соответствующего наибольшему положительному собственному значению матрицы смежности [22]:

Таблица 1

Базовая информация о соавторстве

Число авторов	32434
Число НП, опубликованных в соавторстве	91113
Максимальное число соавторов у НП	17
Среднее число соавторов, приходящихся на одну НП	2,3
Среднее число НП, приходящихся на одного автора	2,8

$$C_E(i) = e_i = \frac{1}{\lambda_{max}} \sum_{j=1}^n a_{ji} \cdot e_j. \quad (10)$$

Здесь λ_{max} — максимальное собственное значение; $\mathbf{e} = (e_1, e_2, \dots, e_n)$ — собственный вектор матрицы смежности, соответствующий λ_{max} .

2. Вычислительный эксперимент. 2.1. *Исходные данные.* Рассматриваемые сети соавторства сконструированы на основе информации, извлеченной из БД RePEc. В работе [12] представлена процедура фильтрации “сырых” данных и определения множества авторов, имеющих публикации в соавторстве. Краткая информация о рассматриваемых данных приведена в табл. 1.

К основным свойствам структуры невзвешенной сети соавторства N^{ca} , исследованной в работе [12], относятся масштабная инвариантность (распределение степеней следует степенному закону) и принадлежность к “малому миру” (небольшое среднее расстояние между узлами [23] и сравнительно высокий локальный коэффициент кластеризации [24]).

Целью данной работы является определение ключевых авторов и выявление зависимости рангов авторов от метода определения их вкладов в публикации и применяемых мер центральности. Предметом изучения являются три сети соавторства:

сеть $N^{ca} = (V^{ca}, E^{ca})$ — невзвешенная сеть соавторства;

сеть $N_T^{ca} = (V^{ca}, E^{ca}, W_T)$, $W_T = (w_{ij}^t)$, вес w_{ij}^t определен T -методом;

сеть $N_F^{ca} = (V^{ca}, E^{ca}, W_F)$, $W_F = (w_{ij}^f)$, вес w_{ij}^f определен F -методом.

Сети N^{ca} , N_T^{ca} и N_F^{ca} имеют одинаковые размеры ($n = |V^{ca}| = 32434$, $m = |E^{ca}| = 73969$) и размеры максимальной компоненты ($|V| = 29270$, $|E| = 71780$, включает 90 % авторов). Далее будем вычислять меры центральности для максимальных компонент сетей соавторства и использовать для них обозначения N^{ca} , N_T^{ca} и N_F^{ca} .

2.2. *Применение мер центральности к сетям соавторства.* Для сетей N^{ca} , N_T^{ca} и N_F^{ca} вычислены четыре меры центральности, приведенные в п. 1.2. В табл. 2–4 для каждой меры центральности приведены 10 авторов, имеющих наивысшие ранги в рамках соответствующих сетей соавторства.

В совокупности в таблицах 2–4 зафиксирован 61 автор, 39 из них входят в 5 % авторов, имеющих максимальное значение h -индекса согласно системе ранжирования IDEAS/RePEc [25], 15 из оставшихся авторов, согласно поисковой системе Google Scholar, имеют h -индекс в диапазоне 5–58 (публикации после 2016 г.). То есть большинство авторов имеют высокий индекс цитирования.

Ранги большинства представленных в табл. 2–4 авторов разнятся в зависимости от способа вычисления вкладов авторов в публикации. Оценим взаимосвязь между одноименными мерами центральности для сетей N^{ca} , N_T^{ca} и N_F^{ca} с использованием коэффициентов корреляции Пирсона r и ранговой корреляции Спирмена ρ . Результаты приведены в табл. 5.

Таблица 2

Авторы сети N^{ca} , имеющие наивысшие ранги

$C_D(i)$	$C_B(i)$	$C_C(i)$	$C_E(i)$
Rangan Gupta	John List	Philippe Aghion	Timo Wollmershäuser
Jacques Thisse	Klaus F. Zimmermann	John List	Wolfgang Nierhaus
Philippe Aghion	Jacques Thisse	Jacques Thisse	Wolfgang Meister
John List	F. G. Schneider	Richard Blundell	Steffen Henzel
Richard Blundell	Philippe Aghion	Klaus F. Zimmermann	Kai Carstensen
Peter C. B. Phillips	Richard Blundell	Thierry A. Verdier	Klaus Wohlrabe
Klaus F. Zimmermann	Ronald MacDonald	Daron Acemoglu	Christian Breuer
Jason Shogren	Rangan Gupta	F. G. Schneider	Steffen Elstner
Ronald MacDonald	Peter C. B. Phillips	Harald Uhlig	Tim Oliver Berg
Peter Nijkamp	Monica Roman	Pierre Cahuc	Christian Grimme

Таблица 3

Авторы сети N_T^{ca} , имеющие наивысшие ранги

$C_D^w(i)$	$C_B^w(i)$	$C_C^w(i)$	$C_E(i)$
Rangan Gupta	Luis Alberiko Gil-Alana	Jan-Egbert Sturm	Timo Wollmershäuser
Jan-Egbert Sturm	Jacques Thisse	Hans-Werner Sinn	Wolfgang Nierhaus
Hans-Werner Sinn	Rangan Gupta	John Hassler	Steffen Henzel
Timo Wollmershäuser	Jan-Egbert Sturm	Gilles Saint-Paul	Wolfgang Meister
Luis Alberiko Gil-Alana	G. M. Caporale	Michael P. Devereux	Kai Carstensen
Michael McAleer	Thierry A. Verdier	Xavier Vives	Christian Breuer
Peter Nijkamp	Yves Zenou	Giancarlo Corsetti	Klaus Wohlrabe
Peter H. Egger	Michael McAleer	Edward John Driffill	Nikolay Hristov
Pierre Pestieau	Jean Tirole	Thierry A. Verdier	Michael Kleemann
Wolfgang Nierhaus	John Gibson	Harold James	Steffen Elstner

Таблица 4

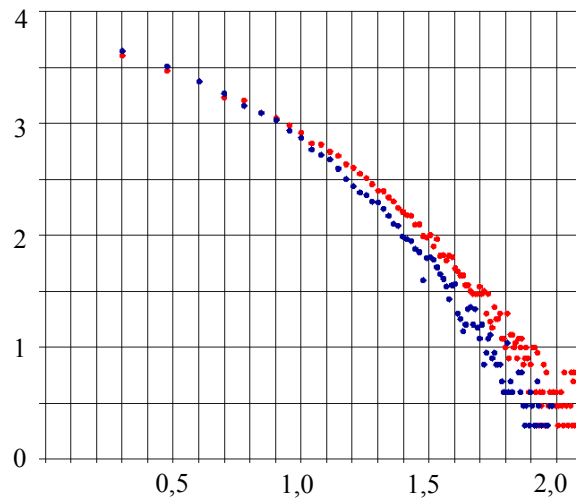
Авторы сети N_T^{ca} , имеющие наивысшие ранги

$C_D^w(i)$	$C_B^w(i)$	$C_C(i)$	$C_E(i)$
Michael McAleer	Jacques Thisse	Jacques Thisse	P. K. Narayan
Rangan Gupta	L. A. Gil-Alana	Thierry A. Verdier	Russell Smyth
Peter Nijkamp	Rangan Gupta	Yves Zenou	Seema Narayan
Luis Alberiko Gil-Alana	Jean Tirole	Jean Tirole	Vinod Mishra
Peter H. Egger	Michael McAleer	Jean-Jacques Laffont	Hooi Hooi Lean
Peter C. B. Phillips	Thierry A. Verdier	A. J.-L. de Palma	Susan Sunila Sharma
Pierre Pestieau	Yves Zenou	Alberto Bisin	Mita Bhattacharya
Jacques Thisse	Jean-Jacques Laffont	Jess Benhabib	Joakim Westerlund
Jean Tirole	A. J.-L. de Palma	Simon Anderson	Ingrid Nielsen
G. M. Caporale	Jess Benhabib	Drew Fudenberg	Anthony Makin

Таблица 5

Корреляция мер центральности сетей N^{ca} , N_T^{ca} , N_F^{ca}

Мера \ Сети	$C_D^w(i)$	$C_B^w(i)$	$C_C^w(i)$	$C_E^w(i)$
$N_T^{ca} - N_F^{ca}$	$r = 0,9353$ $\rho = 0,2948$	$r = 0,9239$ $\rho = 0,5783$	$r = 0,9180$ $\rho = 0,2280$	$r = -0,0007$ $\rho = 0,5780$
$N^{ca} - N_F^{ca}$	$r = 0,8638$ $\rho = 0,4229$	$r = 0,6916$ $\rho = 0,5005$	$r = 0,7764$ $\rho = 0,2208$	$r = -0,0007$ $\rho = -0,0656$
$N^{ca} - N_T^{ca}$	$r = 0,8797$ $\rho = 0,2566$	$r = 0,6957$ $\rho = 0,5135$	$r = 0,8457$ $\rho = 0,2295$	$r = 0,8853$ $\rho = 0,0195$

Рис. 1. Распределения значений центральности узлов сетей N_T^{ca} и N_F^{ca} по степени

На рис. 1 представлены распределения значений центральности узлов сетей N_T^{ca} и N_F^{ca} по степени (красный и синий цвета соответственно, логарифмический масштаб).

Таким образом, в рамках сетей согласно коэффициенту r меры центральности по степени, близости и посредничеству имеют значительную степень корреляции независимо от способа определения весов. Что касается коэффициента ранговой корреляции, то наибольшая близость отмечается для центральности по посредничеству. Если обратиться к табл. 2–4, можно заметить, что состав авторов в колонке C_B^w наиболее близок (7 авторов) для сетей N_T^{ca} и N_F^{ca} , однако ранги половины общих авторов разнятся.

3. Пример. Проиллюстрируем методику вычисления весов ребер и ранжирования узлов согласно мерам центральности на примере одной из компонент невзвешенной сети соавторства N^{ca} (см. [12], табл. 2), состоящей из 12 узлов (r_1, r_2, \dots, r_{12}) и 15 ребер, которую обозначим C_{12} (рис. 2). Список авторов, составляющих множество узлов компоненты, представлен в табл. 6.

Представим схему вычисления весов ребер компоненты C_{12} с применением T - и F -методов на примере фрагмента компоненты C_{12} , включающей авторов r_4, r_5, r_7 . Указанные авторы задекларировали в БД пять публикаций, обозначим их p_1, p_2, \dots, p_5 . На рис. 3 представлен фрагмент сети публикаций (см. п. 1.1), поясняющий соответствие между рассматриваемыми авторами и публикациями.

Схема вычисления весов ребер, $w(i, j)$ — вес ребра (i, j) :

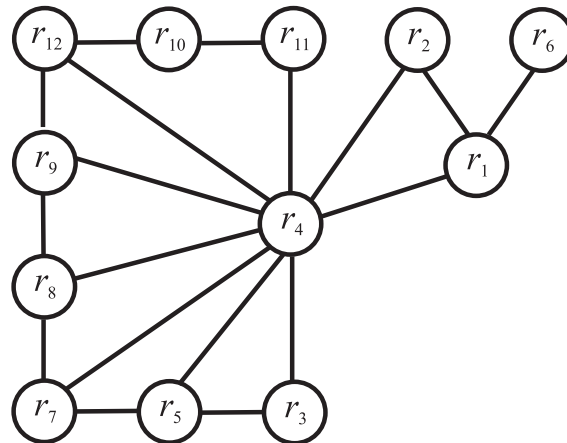
Рис. 2. Компонента C_{12}

Таблица 6

Список авторов компоненты C_{12}

Узел	Идентификатор автора в БД	Автор	Число НП
r_1	4164	Péter Földvári	15
r_2	4165	Bas van Leeuwen	15
r_3	9336	Bas J. P. Van Bavel	1
r_4	9337	Jan Luiten van Zanden	9
r_5	9346	Tine De Moor	5
r_6	9354	Jieli Li	1
r_7	9369	Jaco Zuijderduijn	4
r_8	16760	Maarten R. Prak	1
r_9	17012	Kostadis J. Papaioannou	1
r_{10}	17279	Leticia Arroyo Abad	3
r_{11}	22039	Elwyn Davies	1
r_{12}	30535	Pablo Astorga	1

$T - w(r_4, r_5) = 2$ (две совместных НП: p_1, p_2);

$F - w(r_4, r_5) = 1,5$ (две совместных НП: p_1 , два соавтора; p_2 , три соавтора);

$T - w(r_4, r_7) = 1$ (одна совместная НП: p_2);

$F - w(r_4, r_7) = 0,5$ (одна совместная НП: p_2 , три соавтора);

$T - w(r_5, r_7) = 4$ (четыре совместных НП: p_2, p_3, p_4, p_5);

$F - w(r_5, r_7) = 3,5$ (четыре совместных НП: p_2 , три соавтора; p_3, p_4, p_5 по два соавтора).

На рис. 4 представлены взвешенные варианты рассматриваемого фрагмента компоненты. Веса ребер рассчитаны двумя методами по схеме, представленной выше.

3.1. *Меры центральности узлов C_{12} .* Для узлов компоненты, представленной в трех вариантах: невзвешенным (обозначен UW), взвешенным согласно T -методу (обозначен TW) и взвешенным согласно F -методу (обозначен FW) графами, вычислены меры центральности (п. 1.2) и определены ранги узлов. В табл. 7 представлены ранги узлов, установленные согласно значениям мер C_C (п. 1.2.2), C_B (п. 1.2.3) и C_E (п. 1.2.4). Если несколько узлов имеют одно и то же значение меры центральности, то каждому из них соответствует диапазон рангов, размер которого равен числу таких узлов. Табл. 7 иллюстрирует раз-

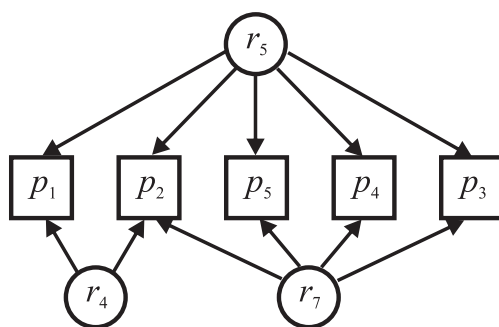


Рис. 3. Фрагмент сети публикаций

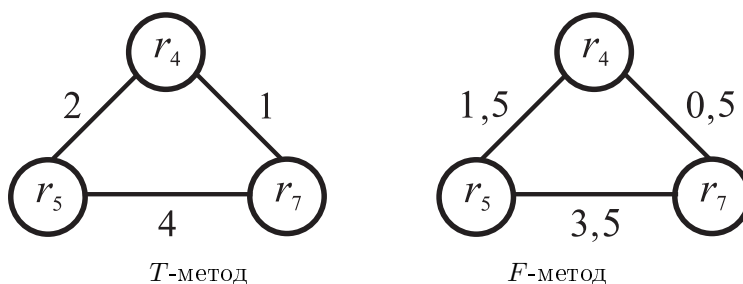
Рис. 4. Взвешенные варианты фрагмента компоненты C_{12}

Таблица 7

Ранги узлов согласно мерам $C_B(\cdot)$, $C_C(\cdot)$, $C_E(\cdot)$

Узел	Ранг согласно $C_C(r_i)$			Ранг согласно $C_B(r_i)$			Ранг согласно $C_E(r_i)$		
	UW	TW	FW	UW	TW	FW	UW	TW	FW
r_1	2-4	4-5	4-5	3-4	4-5	4	2-3	3-4	5-6
r_2	2-4	4-5	4-5	3-4	4-5	5	2-3	3-4	5-6
r_3	8-10	8-10	7-9	5-12	6-12	6-12	9-11	7-9	8-10
r_4	1	1	1	1	1	1	1	1	1
r_5	6-7	3	3	5-12	2-3	2-3	6-7	11	11
r_6	11	12	12	5-12	6-12	6-12	8	10	7
r_7	6-7	6	6	5-12	6-12	6-12	6-7	6	4
r_8	8-10	8-10	7-9	5-12	6-12	6-12	9-11	7-9	8-10
r_9	8-10	8-10	7-9	5-12	6-12	6-12	9-11	7-9	8-10
r_{10}	2-4	2	2	2	2-3	2-3	4	5	3
r_{11}	5	7	11	5-12	6-12	6-12	5	2	2
r_{12}	12	11	10	5-12	6-12	6-12	12	12	12

ницу рангов в зависимости от варианта представления компоненты. Отметим, что только узел r_4 имеет один и тот же ранг независимо от способа определения весов ребер для всех мер центральности.

В табл. 8 представлены ранги узлов (авторов) в соответствии с мерами C_B (обозначена В), C_C (обозначена С), C_E (обозначена Е) для UW-варианта компоненты C_{12} . Ранг помечен знаком "+", если несколько узлов имеют один и тот же ранг. Для каждого узла вычисляется его вес, равный сумме обратных значений рангов. Аналогично вычисляются

Таблица 8

Ранги и веса узлов C12 (UW-представление)

Ранг	r_1			r_2			r_3			r_4			r_5			r_6			r_7			r_8			r_9			r_{10}			r_{11}			r_{12}		
	В	С	Е	В	С	Е	В	В	С	Е	В	С	Е	В	С	Е	В	С	Е	В	С	Е	В	С	Е	В	С	Е	В	С	Е	В	С	Е		
1																																				
2																																				
3																																				
4																																				
5																																				
6																																				
7																																				
8																																				
9																																				
10																																				
11																																				
12																																				
Вес	1,019			1,019			0,329			3,000			0,426			0,334			0,426			0,329			0,329			1,083			0,518			0,285		

Таблица 9

Упорядочивание авторов C_{12} согласно весам (UW-вариант)

Узел	Ранг	Автор
r_4	1	Jan Luiten van Zanden
r_{10}	2	Leticia Arroyo Abad
r_1	3–4	Péter Földvári
r_2	3–4	Bas van Leeuwen
r_{11}	5	Elwyn Davies
r_5	6–7	Tine De Moor
r_7	6–7	Jaco Zuijderduijn
r_6	8	Jieli Li
r_3	9–11	Bas J. P. Van Bavel
r_9	9–11	Kostadis J. Papaioannou
r_8	9–11	Maarten R. Prak
r_{12}	12	Pablo Astorga

Таблица 10

Упорядочивание авторов согласно суммарным весам

Ранг	Автор (узел)	Степень узла согласно варианту			h -индекс	h -индекс с 2016 г.
		UW	TW	FW		
1	Jan Luiten van Zanden	9	11	9	53	36
2	Leticia Arroyo Abad	3	4	3	11	10
3–4	Péter Földvári	3	16	15	16	12
3–4	Bas van Leeuwen	3	16	15	21	17
5	Tine De Moor	2	6	5	24	18
6	Jaco Zuijderduijn	2	5	4	10	8
7	Jieli Li	2	2	1		
8	Elwyn Davies	2	2	1	5	5
9–10	Bas J. P. Van Bavel	1	1	1		
9–10	Kostadis J. Papaioannou	1	1	1		
11	Maarten R. Prak	1	1	1		
12	Pablo Astorga	1	1	1		

веса узлов для взвешенных (TW, FW) вариантов представления компоненты C_{12} . Теперь узлы можно упорядочить согласно весам, т. е. присвоить им новые ранги.

В табл. 9 приведены авторы, упорядоченные по весам соответствующих узлов для UW-представления компоненты C_{12} . Авторам присвоены новые ранги, учитывающие значения трех мер центральности: C_B , C_C и C_E .

В табл. 10 указаны ранги авторов, учитывающие веса, вычисленные на основе значений трех мер центральности (C_B, C_C, C_E) по трем вариантам (UW, TW, FW) представления компоненты C_{12} . Также приведены степени узлов и значения h -индексов авторов (согласно *GoogleScholar*), отдельно вычисленного для работ, опубликованных не раньше 2016 г. Следует заметить, что h -индекс вычислен с учетом НП авторов, большинство из которых не вошло в рассматриваемое множество НП.

Табл. 10 демонстрирует, что лидирующие авторы имеют высокие значения степени сотрудничества и h -индексов.

Заключение. На основе данных, извлеченных из библиографической БД RePEc, построены два варианта взвешенных сетей соавторства — N_T^{ca} , N_F^{ca} . Сети различаются способом вычисления весов ребер, зависящим от подхода к определению вкладов авторов в публикации. В случае применения “полного” (T) метода сила связи двух соавторов определяется как число совместных публикаций, в случае “частичного” (F) для каждой совместной публикации учитывается доля вклада, приходящаяся на каждого соавтора. Для узлов максимальной компоненты сетей N_T^{ca} , N_F^{ca} , а также для невзвешенной сети N^{ca} , рассмотренной в работе [12], вычислены меры центральности по степени, близости, посредничеству и собственному вектору. Приведены списки авторов, занимающих ключевые позиции в рамках рассматриваемых компонент сетей (см. табл. 2–4). Отмечено, что ключевые авторы имеют высокие индексы цитируемости. Проведено сравнение значений одноименных мер центральности на основе коэффициентов корреляции Пирсона и ранговой корреляции Спирмена (см. табл. 5). Значительная корреляция согласно обоим коэффициентам отмечена для центральности по посредничеству, наименьшая — для центральности по собственному вектору, определяющей “престиж” узла сети. Пример на основе компоненты C_{12} иллюстрирует разницу в вычислении весов ребер с применением T - и F -методов и разницу рангов узлов в зависимости от выбора меры центральности и способа назначения весов ребер.

Список литературы

1. EVERETT M. G., BORGATTI S. P. The centrality of groups and classes // J. of Math. Sociology. 1999. V. 23, iss. 3. P. 181–201.
2. BORGATTI S.P. Identifying sets of key players in a social network // Comput. Math. Organiz. Theory. 2006. V. 12. P. 21–34.
3. BOLLEN J., RODRIGUEZ M. A., VAN DE SOMPEL H. Journal status // Scientometrics. 2006. V. 69, iss. 3. P. 669–687.
4. LEYDESDORFF L. Betweenness centrality as an indicator of the interdisciplinarity of scientific journals // J. of the Amer. Soc. for Inform. Sci. and Technol. 2007. V. 58, iss. 9. P. 1303–1319.
5. PENG T-Q. Assortative mixing, preferential attachment, and triadic closure: A longitudinal study of tie-generative mechanisms in journal citation networks // J. of Informetrics. 2015. V. 9, iss. 2. P. 250–262.
6. NEWMAN M. E. J. Scientific collaboration networks. I. Network construction and fundamental results // Phys. Rev. E. 2001. V. 64, iss. 1. 016131.
7. NEWMAN M. E. J. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality // Phys. Rev. E. 2001. V. 64, iss. 1. 016132.
8. NEWMAN M. E. J. Who is the best connected scientist? A study of scientific coauthorship networks // Complex network. Lect. notes in Phys. 2004. V. 650. P. 337–390.
9. YAN E., DING Y. Applying centrality measures to impact analysis // J. of Amer. Soc. for Inform. Sci. and Technol. 2009. V. 60, iss. 10. P. 2107–2118.
10. UDDIN S., HOSSAIN L., ABBASI A., RASMUSSEN K. Trend and efficiency analysis of co-authorship network // Scientometrics. 2012. V. 90. P. 687–699.
11. YOUNGBLOOD M., LAHTI D. A bibliometric analysis of the interdisciplinary field of cultural evolution // Palgrave Communications. 2018. Art. 120.
12. БРЕДИХИН С. В., ЛЯПУНОВ В. М., ЩЕРБАКОВА Н. Г. Структура и параметры невзвешенной сети соавторства на основе данных БД RePEc // Пробл. информ. 2021 (в печати).
13. REPEC. General principles. [Electron. res.]. <http://repec.org/>.

14. PERIANES-RODRIGUEZ A., WALTMAN L., VAN ECK N. J. Constructing bibliometric networks: A comparison between full and fractional counting // *J. of Informetrics*. 2016. V. 10, iss. 4. P. 1178–1195.
15. БРЕДИХИН С. В., ЛЯПУНОВ В. М., ЩЕРБАКОВА Н. Г. Библиометрические сети научных статей и журналов. Новосибирск, ИВМиМГ СО РАН, 2021, 334 с. [Электрон. рес.]. <https://www.elibrary.ru/item.asp?id=45606936>.
16. NIEMINEN J. On centrality in a graph // *Scandinav. J. of Psych.* 1974. V. 15. P. 322–336.
17. BARRAT A., BARTHELEMY M., PASTOR-SATORRAS R., VESPIGNANI A. The architecture of complex weighted networks // *Proc. of the Nation. Acad. of Sci.* 2004. V. 101, iss. 11. -P. 3747–3752.
18. BAVELAS A. Communication patterns in task-oriented groups // *J. of the Acoustical Soc. of Amer.* 1950. V. 22. P. 271–288.
19. BEAUCHAMP M. A. An improved index of centrality // *Behav. Sci.* 1965. V. 10. P. 161–163.
20. ANTHONISSE J. M. The rush in a directed graph // *Technic. rep. BN 9/71*. Amsterdam: Sticing Mathematisch Centrum, 1971.
21. FREEMAN L. C. A set of measures of centrality based upon betweenness // *Sociometry*. 1977. V. 40. P. 35–41.
22. BONACICH P. Factoring and weighting approaches to status scores and clique identification // *J. of Math. Sociol.* 1972. V. 2. P. 113–120.
23. WATTS D. J. Networks, dynamics and the small-world phenomenon // *Amer. J. of Sociol.* 1999. V. 105, iss. 2. P. 493–527.
24. WATTS D. J., STROGATZ S. H. Collective dynamics of 'small-world' networks // *Nature*. 1998. V. 393. P. 440–442.
25. IDEAS : [Electron. res.]. <https://ideas.repec.org/top/top.person.hindex.html>.



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 электрометрии 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 100 работ и трех монографий.

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 100 works and three monographs.



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 20 работ и монографии 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 20 works and monograph in that area.



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

автоматики 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 и соавтор двух монографий. 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 tow monographs. The current research interests lie in the field of bibliometrics: methods of measuring of scientific.

Дата поступления — 13.09.2021