

# COMMUNITY DETECTION IN THE MULTIPLEX NETWORK OF SCIENTIFIC JOURNAL AUTHORS

S. V. Bredikhin, N. G. Scherbakova

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

---

---

DOI: 10.24412/2073-0667-2025-4-11-24

EDN: TWNDJO

The article presents a model of a multiplex network reflecting a real scheme of collaboration of authors of a scientific journal. The initial data are extracted from the XML archive of the journal articles. The model is presented as a two-layer graph, the vertices of which correspond to the authors of the articles, and the edges — to binary relations of co-authorship and citation. The purpose of the work is to identify non-intersecting communities of authors of the network and is achieved in two stages. At the first phase, the network is reduced to the form of an undirected graph  $G_f$  using “flattening algorithm”, this allows to apply the known algorithms of community detection intended for single-layer networks. So, at the second phase, two traditional clustering algorithms, Walktrap and Infomap, based on the “random walk” method are applied to the flattened network. The result of the algorithm is a set of identifiers of nodes included in the community, by which the original multilayer network structure can be restored. The comparison of the algorithms’ results is made using the *rand*, *adjusted\_rand* and *nmi* indices. For  $G_f$ , all indices demonstrate a high level of similarity in the algorithms’ results. The parameter  $\rho_c$ , which characterizes the degree of overlapping of layers at the community level, serves as a characteristic of multilayering. The study of this parameter showed that most communities have a zero value, i.e. the communities consist of vertices of one of the layers. It should be noted that these are the actors of the co-authorship layer, while the citation layer contains 37 % of single nodes. The results of the analysis are presented in the form of tables.

**Key words:** multiplex network, scientific community, co-authorship, citation, modularity, bibliometrics.

## References

1. Barabási A-L., Pósfai M. Network Science. Cambridge Univ. Press. 456 p. ISBN 1107076269.
2. Radicchi F., Castellano C., Cecconi F., Loreto V., Parisi D. Defining and identifying communities in networks // PNAS. 2004. V. 101. P.2658–2663. DOI: 10.1073/pnas.0400054101.
3. Fortunato S. Community detection in graphs // Phys. Rep. 2010. V. 486, iss. 3–5. P. 75–174. DOI: 10.1016/j.physrep.2009.11.002.
4. Distel’ R. Teoriya grafov. Novosibirsk: Izd-vo In-ta matematiki, 2002. 336 s. ISBN 5-86134-101-X.
5. Peel L., Larremore D. B., Clauset A. The ground truth about metadata and community detection in networks // Sci. Adv. 2017. V. 3, iss. 5. e1602548. DOI: 10.1126/scadv.1602548.
6. Newman M. E. J. Modularity and community structure in networks // Proc. Natl. Acad. Sci. USA. 2006. V. 103. P. 8577–8582. DOI: 10.1037/pnas.0601602103.

---

This work was carried out under state contract with ICMMG SB RAS (FWNM-2025-0005).

7. Newman M. E. J., Girvan M. Finding and evaluating community structure in networks // *Phys. Rev. E*. 2004. V. 69. 026113. DOI: 10.1103/PhysRevE.69.026113.
8. Magnani M., Hanteer O., Interdonato R., Rossi L., Tagarelli A. Community detection in multiplex networks // *arXiv: 0911.1824*. DOI: 10.48550/arXiv.1910.07646.
9. Interdonato R., Tagarelli A., Ienco D., Sallaberry A., Poncelet P. Node-centric community detection in multilayer networks with layer-coverage diversification bias // *Proc. of the 8th Conf. on Complex Networks*. 2017. P. 57–66. Springer Intern. Publ., 2017. DOI: 10.48550/arXiv.1704.03441.
10. Jeub L. G. S., Mahoney M. W., Mucha P. J., Porter M. A. A local perspective on community structure in multilayer networks // *Network Sci.* 2017. V. 5, iss.2. P. 144–163. DOI: 10.48550/arXiv.1510.05185.
11. Kim J., Lee J-G. Community detection in multi-layer graphs: A survey // *ACM SIGMOD Record*. 2015. V. 44, iss. 3. P. 37–48. DOI: 10.1145/2854006.2854012.
12. Huang X., Chen D., Ren T., Wang D. A survey of community detection methods in multilayer networks // *Data Mining and Knowledge Discovery*. 2021. V. 35. P. 1–45. DOI: 10.1007/s10618-020-00716-6.
13. Mucha P. J., Richardson T., Macon K., Porter M. A., Onnela J. P. Community structure in time-dependent, multiscale, and multiplex networks // *Science*. 2010. V. 328, iss. 5980. P. 876–878. DOI: 10.1126/science.1184819.
14. De Domenico M., Lancichinetti A., Arenas A., Rosvall M. Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems // *Phys. Review*. 2015. X 5, 011027. DOI: 10.1103/PhysRevX.5.011027.
15. Afsarmanesh N., Magnani M. Finding overlapping communities in multiplex networks // *Proc. of the 2018 Intern. conf. on Social Informatics*, 2018. DOI: 10.48550/arXiv.1602.03746.
16. Bianconi G. *Multilayer networks. Structure and functions*. Oxford. 2018. Online ISBN: 9780191815676.
17. Lancichinetti A., Fortunato S. Consensus clustering in complex networks // *Sci. Rep.* 2012. V. 2. Art. num. 336. DOI: 10.1038/srep00336.
18. Mondragon R. J., Iacovacci J., Bianconi G. Multilink communities of multiplex networks // *arXiv:1706.09011*. DOI: 10.48550/arXiv.1706.09011.
19. De Domenico M., Solé-Ribalta A., Cozzo E., Kivela M., Moreno Y., Porter M. A., Gómez S., Arenas A. Mathematical formulation of multilayer networks // *Phys. Rev.* 2013. X 3. 041022. DOI:10.1103/PhysRevX.3.041022.
20. Bredihin S. V., Shcherbakova N. G. Vzveshennaya mul'tipleksnaya set' avtorov nauchnogo zhurnala // *Probl. inform.* 2025. N 1. S. 45–59. DOI: 10.24412/2073-0667-2025-1-45-59.
21. Bredihin S. V., Shcherbakova N. G. Strukturnye svoystva mul'tipleksnoj seti avtorov nauchnogo zhurnala // *Probl. inform.* 2025. N 2. S. 8–18. DOI: 10.24412/2073-0667-2025-2-5-18.
22. Boccaletti S., Bianconi G., Criado R., del Genio C. I., Gómez-Gardeñes J., Romance M., Sendiña-Nadal I., Wang Z., Zanin M. The structure and dynamics of multilayer networks // *Phys. Rep.* 2014. V. 544, iss. 1. P. 1–122. DOI: 10.1016/j.physrep.2014.07.001.
23. Wagner S., Wagner D. Comparing clusterings — An overview. 2007. DOI: 10.5445/IR/1000011477. <https://publikationen.bibliothek.kit.edu/1000011477>.
24. Collins L. M., Dent C. W. Omega: A general formulation of the Rand index of cluster recovery suitable for non-disjoint solutions // *Multivariate Behav. Res.* 1988. V. 23, iss. 2. P. 231–242. DOI: 10.1207/s15327906mbr2302\_6.
25. Murray G., Carenini G., Ng R. Using the omega index for evaluating abstractive community detection // *Proc. of Workshop on Evaluation Metrics and System Comparison for Automatic Summarization*, Montréal (Canada), 2012. Assoc. for Comput. Linguistics. P. 10–18.
26. Hanteer O., Rossi L. The meaning of dissimilar: An evaluation of various similarity quantification approaches used to evaluate community detection solutions // *Proc. of the IEEE/ACM*

Intern. conf. on Advances in Social Networks Analysis and Mining, Vancouver (Canada), 2019. P. 513–518. DOI: 10.1145/3341161.3342941.

27. Berlingerio M., Coscia M., Giannotti F. Finding and characterizing communities in multidimensional networks // Intern. conf. on Advances in Social Networks Analysis and Mining (ASONAM). P. 490–494. IEEE Computer Society Washington, DC, USA, 2011. DOI: 10.1107/ASONAM.2011.104.

28. Kim J., Lee J.-G., Lim S. Differential flattening: A novel framework for community detection in multi-layer graphs // ACM Trans. on Intell. Syst. and Technol. (TIST). 2016. V. 8, iss. 2. P. 27:1–27:23. DOI: 10.1145/2898362.

29. De Domenico M., Nicosia V., Arenas A., Latora V. Structural reducibility of multilayer networks // Nature Commun. 2015. V. 6. 6864. DOI: 10.1038/ncomms7864.

30. Bianconi G. Statistical mechanics of multiplex networks: entropy and overlap // Phys. Rev. E. 2013. V. 87, iss. 6. 062806. DOI: 10.1103/PhysRevE.87.062806.

31. Pons P., Latapy M. Computing communities in large networks using random walks. 2006. arXiv: physics/0512106. DOI: 10.48550/arXiv.physics/0512106.

32. Rosvall M., Axelsson D., Bergstrom C. T. Map equation. // Eur. Phys. J. 2009. V. 178. P. 13–23. DOI: 10.1140/epjst/e2010-01179-1.

# ВЫЯВЛЕНИЕ СООБЩЕСТВ В МУЛЬТИПЛЕКСНОЙ СЕТИ АВТОРОВ НАУЧНОГО ЖУРНАЛА

С. В. Бредихин, Н. Г. Щербакова

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

УДК 519.177

DOI: 10.24412/2073-0667-2025-4-11-24

EDN: TWNDJO

Представлена модель мультиплексной сети, отражающая реальную схему сотрудничества авторов научного журнала. Исходные данные извлечены из XML-архива статей журнала. Модель выполнена в виде двухслойного графа, вершины которого соответствуют авторам статей, а ребра — бинарным отношениям соавторства и цитирования. Цель работы состоит в выявлении непересекающихся сообществ авторов сети и достигается в два этапа. На первом этапе сеть приводится к виду неориентированного графа, на втором к построенному графу применяются два традиционных алгоритма кластеризации, основанные на методе случайного блуждания. Выполнен вычислительный эксперимент.

**Ключевые слова:** мультиплексная сеть, научное сообщество, соавторство, цитирование, модульность, библиометрия.

**Введение.** Выявление сетевых сообществ — важный шаг современного сетевого анализа, результаты которого позволяют строить обоснованные предположения о топологии изучаемого сетевого объекта и измерять его параметры. Термин *сетевое сообщество* (*кластер*) не имеет общепринятого определения и интуитивно определяется как группа узлов (*акторов*), которые более плотно связаны друг с другом, чем с другими узлами в сети. Определения сообществ опираются на следующие гипотезы А.-Л. Барабаши [1]. Фундаментальная гипотеза: «*Структура сообщества однозначно закодирована в схеме соединения ее узлов*». Гипотеза связности: «*Сообщество соответствует связному подграфу*». Гипотеза плотности: «*Сообщество является локально плотным подграфом*». В этом контексте полный подграф (клика) является сообществом.

Другой способ определения опирается на разделение степеней каждого узла, принадлежащего сообществу, на внутреннюю степень, означающую число ребер с узлами того же сообщества, и внешнюю, означающую число ребер с узлами других сообществ. При этом сообщество называют *сильным*, если каждый узел имеет большее число ребер внутри сообщества, нежели вне его. Сообщество называют *слабым*, если его общая внутренняя степень превышает его общую внешнюю степень [2]. Таким образом, каждое сильное сообщество является слабым, обратное неверно.

Задача выявления сообществ весьма трудоемка и на сегодняшний день не имеет универсального решения. В обзоре [3] представлены определения основных элементов проблемы

---

Исследования выполнены в рамках государственного задания ИВМиМГ СО РАН (FWNM-2025-0005).

и ряд разработанных методов их решения. Один из традиционных методов выявления сообществ — разбиение сети на заранее определенное число групп приблизительно одинакового размера, так что число ребер между группами минимально. В случае, когда число групп равно двум, процедура кластеризации называется *разрезом*. Основополагающая теорема Менгера (доказана в 1927 г.) о минимальном разрезе, характеризующая связность неориентированного либо ориентированного графа, утверждает, что минимальное число ребер, удаление которых разъединяет узлы  $i$  и  $j$ , равно максимальному числу попарно непересекающихся путей из  $i$  в  $j$  [4].

Учитывая актуальность проблемы, крайне важно построить эффективные процедуры и выбрать соответствующие алгоритмы для идентификации структуры сетевых сообществ. В большинстве случаев сообщества определяются алгоритмически, т. е. они являются конечным продуктом алгоритма, без формального определения [3]. Производительность алгоритмов обнаружения сообществ зачастую оценивают по их способности находить так называемые истинные сообщества. Это широко применяется в синтетических сетях, тогда как в реальных сетях заранее известных сообществ нет. На практике принято обрабатывать дополнительную информацию об узлах сети — *метаданные*. Этот прием следует использовать осторожно, поскольку не существует принятого определения того, каким должен быть правильный результат [5]. Популярной функцией определения качества разбиения на сообщества является *модульность* [6]. В ее основе лежит гипотеза о том, что случайный граф не имеет кластерной структуры. Существование кластеров выявляется путем сравнения фактической плотности ребер в кластере и плотности, которой можно было бы ожидать, если бы вершины графа были связаны независимо от структуры сообществ. Модульность также может выступать в качестве оптимизирующей функции алгоритма кластеризации [7].

Мультиплексная сеть — это модель, используемая для представления реальных объектов с несколькими типами отношений между акторами. Каждый тип отношений представлен отдельным слоем, моделируемым графом. Традиционные методы обнаружения сообществ для обычных графов недостаточны для сложной мультиплексной модели по следующим причинам [8]. Во-первых, без возможности анализа подмножеств слоев некоторые сообщества могут оказаться скрытыми из-за рассмотрения слоев, не имеющих к ним отношения. Во-вторых, алгоритмы, не представляющие различные слои явно, не могут различать различные типы мультиплексных сообществ, например присутствующие только в одном слое или состоящие из определенных комбинаций слоев. В-третьих, без концепции слоя невозможно включить один и тот же узел в разные сообщества в зависимости от слоя, на котором узел активен. Для устранения вышеуказанных ограничений был разработан ряд алгоритмов обнаружения сообществ для мультиплексных сетей, основанных на различных определениях сообщества и различных вычислительных методах. В мультиплексных сетях, где узлы связаны несколькими типами бинарных отношений, возникли новые проблемы в обнаружении сообществ. Современный сетевой анализ появился как перспективный подход к исследованию подобных сетей.

Результатом алгоритмов выявления сообществ в мультиплексных сетях является множество кластеров, каждый из которых содержит непустое подмножество множества узлов. При этом кластеры могут охватывать несколько слоев. Для выявления многослойной природы сообществ были разработаны различные методы. В работе [8] рассматриваются три уровня классификации методов. Верхний уровень различает *глобальные* (выявление всех сообществ в исследуемой сети) или *локальные* (одно сообщество вокруг одного или

нескольких выбранных узлов) методы (см. [9, 10]). Второй уровень связан со *способом*, которым алгоритмы обрабатывают наличие нескольких слоев: *сглаживание (flattening)* — сведение к одному слою; *послойная кластеризация* — обработка каждого слоя независимо для последующего объединения результатов обработки; *многослойная кластеризация* — рассмотрение всех слоев одновременно (см. обзоры [8, 11, 12]). Последний уровень классификации группирует алгоритмы, основанные на конкретных методах, таких как оптимизация целевой функции [13], случайное блуждание [14] или выявление плотных подграфов [15]. Классификация конкретных методов многослойной кластеризации приведена в работе [16], при этом рассматриваются пять основных классов: максимизация многослойной модульности [13]; консенсусная кластеризация [17]; кластеризация на основе свойств многослойных случайных блужданий [14]; кластеризация на основе мультилинков [18]; методы, основанные на тензорном разложении и соответствующих вычислениях [19].

В данной статье представлена двухслойная мультиплексная сеть  $\mathcal{M}^{cc}$ , моделирующая реальную схему сотрудничества авторов научного журнала. Произведена кластеризация узлов сети с использованием метода сглаживания. Апробированы два алгоритма выявления сообществ, основанные на методе случайного блуждания. Выполнена оценка качества разбиения. Данная работа является продолжением работ [20, 21].

**1. Мультиплексная сеть.** Многослойная сеть [22] — это пара  $\mathcal{M} = (G, C)$ , в которой  $G = \{G_\alpha, \alpha \in \{1, \dots, M\}\}$  — семейство ориентированных (либо неориентированных), взвешенных (либо невзвешенных) графов  $G_\alpha = (V_\alpha, E_\alpha)$ , называемых слоями, а  $C = \{E_{\alpha\beta} \subseteq V_\alpha \times V_\beta; \alpha, \beta \in \{1, \dots, M\}, \alpha \neq \beta\}$  — множество взаимосвязей между узлами различных слоев  $G_\alpha$  и  $G_\beta$ ,  $\alpha \neq \beta$ . Элементы  $E_\alpha$  называются *внутрислойными* ребрами, а элементы  $E_{\alpha\beta}$  — *межслойными*.

Мультиплексная сеть  $\mathcal{M}$  является особым видом многослойной сети, в которой  $V_1 = V_2 = \dots = V_M = V$  и межслойные ребра разрешены только между репликами одних и тех же физических узлов, т. е.  $\forall \alpha, \beta \in \{1, \dots, M\}, \alpha \neq \beta (E_{\alpha\beta} = \{(v, v), v \in V\})$ .

*Проекцией* мультиплексной сети называют граф  $proj(\mathcal{M}) = (V_M, E_M)$ , элементы матрицы смежности которого определяются следующим образом:

$$a_{ij} = \begin{cases} 1, & \text{если } a_{ij}^\alpha = 1, 0 \leq \alpha \leq M \\ 0 & \text{в противном случае.} \end{cases} \quad (1)$$

В статье многослойная сеть определяется кортежем  $(A, L, V, E)$ , где  $A$  — множество акторов;  $L = \{l_1, l_2, \dots, l_m\}$  — множество слоев;  $(V, E)$  — граф на  $V \subseteq A \times L$ . Это определение не требует, чтобы все акторы присутствовали во всех слоях. В мультиплексной сети множество  $E$  ограничено внутриуровневыми ребрами, т. е. ребро  $((v_1, l_1), (v_2, l_2)), l_1, l_2 \in L$  существует только если  $l_1 = l_2$ , а множество ребер  $E = \{E_1, E_2, \dots, E_m\}$ .

**2. Сетевые сообщества.** С точки зрения покрытия различают *полную (total)* кластеризацию (каждый узел принадлежит по крайней мере одному сообществу) и *частичную (partial)* кластеризацию (не все узлы распределены по сообществам). Кластеризацией с *перекрываем узлов (node-overlapping)* называется кластеризация, в которой существует по крайней мере один узел, принадлежащий более чем одному кластеру, в противном случае это кластеризация *без перекрывания узлов (node-disjoined)*. Аналогично, если есть по крайней мере один актор, принадлежащий более чем одному кластеру, кластеризация называется кластеризацией с *перекрываем акторов (actor-overlapping)*, в противном случае кластеризацией *без перекрывания акторов (actor-disjoined)*. Заметим, что кластеризация с перекрыванием узлов является также кластеризацией с перекрыванием акторов, однако кла-

стеризация с перекрытием акторов не обязательно является кластеризацией с перекрытием узлов. Наконец, мультиплексное сообщество называется *полуколонной* (*semi-pillar*) на слоях  $L' \subseteq L$ , если для каждого актора  $a \in A$  в сообществе все узлы  $\{(a, l) \in V; l \in L'\}$  входят в сообщество. Если  $L' = L$ , то это *колонна*. Два сообщества-колонны либо не пересекаются, либо пересекаются как по акторам, так и по узлам.

2.1. *Качество кластеризации.* Проверка качества кластеризации — нетривиальная задача даже для однослойных сетей, более того, не существует единого общепринятого определения того, каким должен быть правильный результат. Так, в работе [5] показано, что метаданные (например, аффилиация авторов) — это не то же самое, что «абсолютная истина», и что обращение с ними как с таковыми вызывает серьезные теоретические и практические проблемы. Обнаруженные сообщества и метаданные фиксируют различные аспекты структуры сети. В работе [8] подчеркнута, что в случае мультиплексной сети может оказаться, что не все слои вносят положительный вклад в определение структуры сообществ. А тот факт, что большинство методов обнаружения сообществ всегда дают выход независимо от того, какие слои включены во входную мультиплексную сеть, делает включение большего количества входных слоев потенциально проблематичным. Слои, помимо того что определяются внутренней топологией, также определяются внутренней логикой, которая может быть согласованной или не согласованной с логикой других слоев.

2.2. *Сравнение результатов кластеризации.* Для измерения парного сходства между результатами глобальной кластеризации могут использоваться традиционные критерии оценки разбиений, такие как индексы *Rand*, *Adjusted\_Rand* или *NMI* [23]. Однако эти критерии применимы только в случае неперекрывающихся сообществ. Индекс *Omega* [24] является мерой, которую можно применять для сравнения результатов разбиения в ситуациях, когда оба результата разбиения содержат перекрывающиеся сообщества, один содержит или ни один не содержит [25, 26].

2.3. *Метод сглаживания (flattening)* [27, 28]. В общем случае метод регламентирует процесс формального представления взаимодействия акторов в обозримом формате, например табличном. В нашем случае применяется с целью преобразования многослойной сети в однослойную сеть (возможно, взвешенную) путем игнорирования информации, характерной для индивидуальных слоев, т. е. не принимая во внимание информацию о типе ребер. Его применение позволяет использовать традиционные алгоритмы обнаружения сообществ.

Метод *полного сглаживания* позволяет конструировать невзвешенные графы, где два узла смежны, если их соответствующие акторы смежны в любом из входных слоев, т. е. для мультиплекса это построение проекции сети (1). Преимущество этого метода заключается в том, что существует большое число хорошо апробированных алгоритмов кластеризации для невзвешенных графов, а при наличии взвешенных ребер зачастую подразумевается дополнительный уровень сложности. В свою очередь, *взвешенное сглаживание* отражает важные структурные свойства исходной мультиплексной сети в форме весов, назначенных ребрам. Например, один из способов назначения веса, в котором параметр вес означает число слоев, в которых присутствует данное ребро:  $w(u, v) = |\{i, (u, v) \in E_i\}|$ . В работах [27, 28] приведены также другие способы определения весов. Альтернативный подход к полному сглаживанию, состоящий из компромисса между оригинальной многослойной сетью и сетью, полученной полным сглаживанием, приведен в работе [29].

Методы сглаживания приводят к образованию сообществ-колонн, так как акторы из разных слоев представлены одним узлом в сглаженном графе (далее —  $G_f$ ). В результа-

те работы алгоритма могут возникнуть сообщества, которые не являются сообществом в каком-либо отдельном слое. Сложность алгоритма в данном случае соответствует сложности шага обнаружения сообществ. Более высокое сходство слоев, например с точки зрения присутствия ребер между одноименными акторами, может повлечь меньшее время выполнения однослойного алгоритма.

**2.4. Оценка результатов кластеризации.** В однослойной сети плотность сообщества определяется соотношением между числом существующих ребер среди узлов сообщества и числом всех возможных ребер. В многослойной сети плотность может обеспечиваться одним слоем, а может — несколькими. В работе [27] определен параметр  $\rho_c$ , характеризующий сообщества многослойной сети. Пусть  $c$  — сообщество, сформированное алгоритмом поиска сообществ в многослойной сети. Обозначим  $M$  — множество слоев;  $P$  — множество пар  $(u, v)$ , связанных по крайней мере в одном слое сети;  $P^2$  — множество пар, связанных по крайней мере в двух слоях;  $P_c$  — подмножество пар из  $P$ , входящих в сообщество  $c$ . Рассмотрим подмножество пар  $P_c^2 \subseteq P^2$ , состоящее из пар множества  $P^2$ , входящих в  $c$ . Обозначим ребро, принадлежащее слою  $m \in M$ , триадой  $(u, v, m)$ ;  $E$  — множество всех ребер исходной сети.

Определяется мера (2), названная *избыточностью*, отражающая ситуацию, при которой множество узлов, составляющих сообщество в одном слое, составляет сообщество также и в других слоях. Эта мера может рассматриваться как индикатор избыточности ребер: чем больше слоев соединяют каждую пару узлов в сообществе, тем выше будет избыточность. Теперь можно определить *нормализованную* избыточность, подсчитав число пар, имеющих избыточные связи, нормализуя по теоретическому максимуму:

$$\rho_c = \sum_{(u, v) \in P_c^2} \frac{|m : \exists (u, v, m) \in E|}{|M| \cdot |P_c|}. \quad (2)$$

Параметр  $\rho_c$  принимает значения в диапазоне  $[0, 1]$ . Поскольку для сообществ, состоящих из одного элемента,  $|\rho_c| = 0$ , такие сообщества не рассматриваются. По существу, параметр  $\rho_c$  подобен мере перекрытия слоев на уровне сообществ [30].

### 3. Вычислительный эксперимент.

**3.1. Данные.** Взвешенная двухслойная мультиплексная сеть авторства  $M^{cc}$  построена на основе данных, извлеченных из записей электронного XML-архива научного журнала «Сахарный диабет» (ISSN 2072-0378), находящегося в свободном доступе. Рассматривается множество статей  $P$ , опубликованных в журнале,  $|P| = 991$ , и множество авторов  $A$ ,  $|A| = 1694$ , участвовавших в их создании. В свою очередь,  $P$  представлено объединением множеств  $P^m$  — статей, созданных в соавторстве, и множества  $P^s$  — индивидуальных статей (в нашем случае  $|P^m| = 843$ ,  $|P^s| = 148$ ). Аналогично, множество авторов  $A$  представлено объединением множеств  $A = A^m \cup A^s$ , где  $A^m$  — подмножество авторов, имеющих совместные публикации,  $A^s$  — имеющих исключительно индивидуальные публикации (в нашем случае  $|A^m| = 1676$ ,  $|A^s| = 18$ ). Для вычислений использовался пакет *igraph* на основе *R*.

В качестве узлов сети выступают авторы статей. Первый слой ( $l_1$ ) — граф соавторства  $G_1 = (V, E_1)$ ,  $|V| = |A|$ , узлы соответствуют авторам; ребро  $(u, v, 1)$  между двумя узлами устанавливается в случае, если соответствующие авторы имеют совместные публикации; вес ребра равен числу публикаций. Второй слой ( $l_2$ ) — ориентированный граф цитирования  $G_2 = (V, E_2)$ ; ребро  $(u, v, 2)$  между цитирующим и цитируемым устанавливается,

Таблица 1

Параметры слоев сети  $\mathcal{M}^{cc}$  и графа  $G_f$ 

Граф	$n$	$m$	$nc$	$mcs$	$dens$	$cc$	$apl$	$dia$
$G_1$	1676	16418	96	1112	0,005848324	0,3627196	3,781167	11
$G_2$	1075	21171	763	305	0,018337014	0,2198572	2,815648	7
$G_f$	1682	24134	49	1484	0,017071126	0,2437155	2,607897	7

если в списке литературы статьи, в которой участвует цитирующий автор, присутствует ссылка на статью из журнала, в которой участвует цитируемый автор. Детали построения приведены в работе [20]. В качестве сглаженной сети рассматривается невзвешенный граф (сглаживания)  $G_f = (V, E)$ , в котором два узла смежны, если их соответствующие акторы смежны хотя бы в одном слое, ориентация второго слоя не учитывается.

Параметры слоев и графа  $G_f$  приведены в табл. 1. Обозначения:  $n$  — число узлов;  $m$  — число ребер;  $nc$  — число компонент;  $mcs$  — размер максимальной компоненты;  $dens$  — реберная плотность;  $cc$  — коэффициент кластеризации (глобальный — отношение числа треугольников к числу связных троек);  $apl$  — средняя длина пути;  $dia$  — диаметр. Одиночные узлы не включены в состав слоев сети.

3.2. *Алгоритмы.* Рассмотрим два алгоритма, использующие методы, основанные на случайных блужданиях. Они предполагают, что актер, случайным образом следующий по ребрам в сети, будет иметь тенденцию попадать в ловушку внутри сообществ из-за более высокой плотности ребер между узлами внутри одного сообщества, реже перемещаясь из одного сообщества в другое. Результатом работы алгоритмов является множество непересекающихся сообществ, при этом каждый актер попадает в какое-либо сообщество.

3.2.1. Алгоритм *Walktrap*, представленный в работе [31], выявляет плотные подграфы исходного графа на основе случайного блуждания. Агломеративный алгоритм объединяет узлы в кластеры на основе определенной в работе меры сходства (расстояния) между узлами. На начальном этапе каждый узел является кластером. Далее, по критерию расстояния, на каждом этапе два соседних кластера объединяются в новый кластер, образуя новое разбиение узлов. Пересчитываются расстояния. Ключевой момент алгоритма — способ выбора сообществ для слияния и тот факт, что расстояния можно эффективно обновлять. Сложность алгоритма оценивается  $O(mn^2)$ , где  $n$  и  $m$  — число узлов и ребер соответственно.

Выбор результирующего разбиения производится на основе модульности. Модульность  $Q$  — это функция качества, которая сравнивает число ребер между узлами сообществ с числом, которое можно было бы ожидать при группировании случайным образом. То есть для идентификации сообществ в сети требуется найти разбиение, которое максимизирует модульность. Если вероятность того, что два узла связаны ребром в графе, определяется равенством  $p_{ij} = \frac{k_i \cdot k_j}{\langle k \rangle \cdot N}$ , а  $c_i$  — это сообщество, к которому принадлежит узел  $i$ , то параметр модульность  $Q$  определяется равенством

$$Q = \frac{1}{\langle k \rangle \cdot N} \sum_{ij} (a_{ij} - p_{ij}) \cdot \delta(c_i, c_j), \quad (3)$$

где  $\delta$  — дельта Кронекера;  $a_{ij}$  — элемент матрицы смежности графа;  $k_i$  — степень узла  $i$ ;  $\langle k \rangle$  — среднее значение степени узлов [6].

Таблица 2

Статистика структуры сообществ

Граф \ Параметры	$K_1$	$K_2$	$K_3$	$K_4$	$Q$
Алгоритм <i>Walktrap</i>					
$G_f$	230	552	4	46	0,2873692
$G_1$	259	305	3	15	0,5499898
$G_2$	189	433	2	143	0,1195941
Алгоритм <i>Infomap</i>					
$G_f$	146	458	8	0	0,3215479
$G_1$	211	123	3	0	0,5809985
$G_2$	16	969	0	0	0,0218763

Таблица 3

Индексы сравнения результатов разбиений графов с использованием алгоритмов *Walktrap* и *Infomap*

Граф \ Индекс	$G_f$	$G_1$	$G_2$
<i>rand</i>	0,9475448	0,9686040	0,3534814
<i>adjustedrand</i>	0,7048701	0,3995592	0,0837723
<i>nmi</i>	0,8706313	0,8995698	0,2737676

3.2.2. Алгоритм *Infomap* [32] идентифицирует составляющие сеть модули путем эффективного описания процесса распространения информации по сети. Основан на идее случайного блуждания по сети с целью минимизировать длину пути обхода узлов, что позволяет выявлять сообщества на основе сетевых потоков информации. Выделяется группа узлов, в которой обмен информацией осуществляется достаточно быстро, узлы этой группы объединяются, далее она выступает как единый модуль. Связи между модулями, построенными подобным образом, отражают сетевые информационные потоки. Для описания этого процесса применяется эффективный код Хаффмана, позволяющий генерировать уникальные имена узлов. Этот прием позволяет построить карту сети, в результате случайного блуждания (путешествия) по которой можно выделить модули, в которых «путешественник» провел большую часть времени. Вычислительная сложность алгоритма  $O(n \log n)$ .

3.3. *Результаты разбиения.* Для сравнения приводятся результаты разбиения на сообщества для графа  $G_f$ , а также для слоев, представленных графами  $G_1$  и  $G_2$ . Результаты работы алгоритмов *Walktrap* и *Infomap* представлены в табл. 2.

В заголовках табл. 2 используются следующие сокращения:  $K_1$  — число кластеров;  $K_2$  — размер максимального кластера;  $K_3$  — число кластеров с числом узлов меньше максимального и больше 30;  $K_4$  — число одиночных кластеров.

При разбиении на сообщества взвешенного ориентированного графа  $G_2$  с помощью алгоритма *Infomap* приблизительно 90 % узлов попало в максимальное сообщество. Из 458 узлов, входящих в максимальное сообщество при разбиении графа сглаживания  $G_f$ ,

Таблица 4  
Значения параметра  $\rho_c$

<i>Infomap</i>			<i>Walktrap</i>		
$c(\cdot)$	$n$	$\rho_c$	$c(\cdot)$	$n$	$\rho_c$
106	2	1,000	26	2	1,000
55	10	0,756	63	2	1,000
63	10	0,600	14	10	0,756
11	7	0,500	42	7	0,583
72	9	0,438	20	8	0,529
22	26	0,244	61	5	0,400
6	10	0,211	58	15	0,365
2	69	0,202	19	7	0,357
73	11	0,185	12	17	0,239
29	12	0,179	35	9	0,222
38	33	0,166	7	62	0,221
35	34	0,163	6	69	0,202
25	26	0,160	9	32	0,167
1	458	0,152	18	12	0,158
5	31	0,113	1	552	0,147
10	51	0,087	5	74	0,083
59	31	0,078	53	13	0,014
32	10	0,077			
9	31	0,068			

114 узлов попало в максимальное сообщество при кластеризации  $G_1$  (из 123), и 420 входят в максимальное при кластеризации  $G_2$  (из 969).

Сравнение результатов разбиений с помощью алгоритмов *Walktrap* и *Infomap* представлено в табл. 3. Наиболее стабильные результаты у графа сглаживания  $G_f$ .

3.4. *Характеристика сообществ на основе параметра  $\rho_c$* . В табл. 4 приведены ненулевые значения параметра  $\rho_c$ , характеризующего сообщества, выявленные с помощью алгоритмов *Infomap* и *Walktrap*. Здесь  $c(\cdot)$  — номер сообщества,  $n$  — число узлов в сообществе. Для алгоритма *Infomap* ненулевое значение  $\rho_c$  имеют приблизительно 13 % кластеров, а для алгоритма *Walktrap* — 7,4 % кластеров. При этом в обоих случаях нулевое значение имеют в основном кластеры небольшого размера. Следует также заметить, что большое число нулевых значений объясняется тем, что в слое цитирования 37 % акторов представлены одиночными вершинами.

Отметим, что кластеры наибольшего размера (458 и 552), полученные с использованием алгоритмов *Infomap* и *Walktrap*, пересекаются по 436 узлам.

**Заключение.** Рассмотрен один из методов выявления непересекающихся сообществ многослойной сети в применении к двухслойной мультиплексной сети авторов научного журнала. Он осуществляет построение отображения многослойной сети на однослойную с применением метода сглаживания. Это позволило использовать известные алгоритмы выявления сообществ, предназначенные для однослойных сетей. Результатом работы алгоритма является множество идентификаторов узлов, входящих в сообщество, по которым можно восстановить исходную многослойную сетевую структуру. В качестве характеристики многослойности выступает параметр  $\rho_c$ , характеризующий степень перекрытия сло-

ев на уровне сообществ. Рассмотрены два алгоритма выявления сообществ, основанные на технике свободного блуждания — *Walktrap* и *Infomap*. Сравнение результатов работы алгоритмов проведено с использованием индексов *rand*, *adjusted\_rand* и *nmi*. Для  $G_f$  все индексы демонстрируют высокий уровень сходства результатов работы алгоритмов. Использование параметра  $\rho_c$  показало, что большинство сообществ имеют нулевое значение параметра, т. е. сообщества состоят из вершин одного из слоев. В дальнейшем предполагается апробировать методы послойной и многослойной кластеризации к анализируемой сети.

Авторы благодарят А. В. Феофанова за техническую поддержку на этапе выполнения вычислительного эксперимента.

## Список литературы

1. Barabási A-L., Pósfai M. Network Science. Cambridge Univ. Press. 456 p. ISBN 1107076269.
2. Radicchi F., Castellano C., Cecconi F., Loreto V., Parisi D. Defining and identifying communities in networks // PNAS. 2004. V. 101. P. 2658–2663. DOI: 10.1073/pnas.0400054101.
3. Fortunato S. Community detection in graphs // Phys. Rep. 2010. V. 486, iss. 3–5. P. 75–174. DOI: 10.1016/j.physrep.2009.11.002.
4. Дистель Р. Теория графов. Новосибирск: Изд-во Ин-та математики, 2002. 336 с. ISBN 5-86134-101-X.
5. Peel L., Larremore D. B., Clauset A. The ground truth about metadata and community detection in networks // Sci. Adv. 2017. V. 3, iss. 5. e1602548. DOI: 10.1126/scadv.1602548.
6. Newman M. E. J. Modularity and community structure in networks // Proc. Natl. Acad. Sci. USA. 2006. V. 103. P. 8577–8582. DOI: 10.1037/pnas.0601602103.
7. Newman M. E. J., Girvan M. Finding and evaluating community structure in networks // Phys. Rev. E. 2004. V. 69. 026113. DOI: 10.1103/PhysRevE.69. 026113.
8. Magnani M., Hanteer O., Interdonato R., Rossi L., Tagarelli A. Community Detection in Multiplex Networks // arXiv: 0911.1824. DOI:10.48550/arxiv: 0911.1824.
9. Interdonato R., Tagarelli A., Ienco D., Sallaberry A., Poncelet P. Node-centric community detection in multilayer networks with layer-coverage diversification bias // Proc. of the 8th Conf. on Complex Networks. 2017. P. 57–66. Springer Intern. Publ., 2017. DOI: 10.48550/arXiv.1704.03441.
10. Jeub L. G. S., Mahoney M. W., Mucha P. J., Porter M. A. A local perspective on community structure in multilayer networks // Network Sci. 2017. V. 5, iss. 2. P. 144–163. DOI: 10.48550/arXiv.1510.05185.
11. Kim J., Lee J-G. Community detection in multi-layer graphs: A survey // ACM SIGMOD Record. 2015. V. 44, iss. 3. P. 37–48. DOI: 10.1145/2854006.2854012.
12. Huang X., Chen D., Ren T., Wang D. A survey of community detection methods in multilayer networks // Data Mining and Knowledge Discovery. 2021. V. 35. P. 1–45. DOI: 10.1007/s10618-020-00716-6.
13. Mucha P. J., Richardson T., Macon K., Porter M. A., Onnela J. P. Community structure in time-dependent, multiscale, and multiplex networks // Science. 2010. V. 328, iss. 5980. P. 876–878. DOI: 10.1126/science.1184819.
14. De Domenico M., Lancichinetti A., Arenas A., Rosvall M. Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems // Phys. Review. 2015. X 5, 011027. DOI: 10.1103/PhysRevX.5. 011027.

15. Afsarmanesh N., Magnani M. Finding overlapping communities in multiplex networks // Proc. of the 2018 Intern. conf. on Social Informatics, 2018. DOI: 10.48550/arXiv.1602.03746.
16. Bianconi G. Multilayer networks. Structure and functions. Oxford. 2018. Online ISBN: 9780191815676.
17. Lancichinetti A., Fortunato S. Consensus clustering in complex networks // Sci. Rep. 2012. V. 2. Art. num. 336. DOI: 10.1038/srep00336.
18. Mondragon R. J., Iacovacci J., Bianconi G. Multilink communities of multiplex networks // arXiv:1706.09011. DOI: 10.48550/arXiv.1706.09011.
19. De Domenico M., Solé-Ribalta A., Cozzo E., Kivelä M., Moreno Y., Porter M. A., Gómez S., Arenas A. Mathematical formulation of multilayer networks // Phys. Rev. 2013. X 3. 041022. DOI:10.1103/PhysRevX.3.041022.
20. Бредихин С. В., Щербакова Н. Г. Взвешенная мультиплексная сеть авторов научного журнала // Пробл. информ. 2025. № 1. С. 45–59. DOI: 10.24412/2073-0667-2025-1-45-59.
21. Бредихин С. В., Щербакова Н. Г. Структурные свойства мультиплексной сети авторов научного журнала // Пробл. информ. 2025. № 2. С. 8–18. DOI: 10.24412/2073-0667-2025-2-5-18.
22. Vocaletti S., Bianconi G., Criado R., del Genio C. I., Gómez-Gardeñes J., Romance M., Sendiña-Nadal I., Wang Z., Zanin M. The structure and dynamics of multilayer networks // Phys. Rep. 2014. V. 544, iss. 1. P. 1–122. DOI: 10.1016/j.physrep.2014.07.001.
23. Wagner S., Wagner D. Comparing clusterings — An overview. 2007. DOI: 10.5445/IR/1000011477. <https://publikationen.bibliothek.kit.edu/1000011477>.
24. Collins L. M., Dent C. W. Omega: A general formulation of the Rand index of cluster recovery suitable for non-disjoint solutions // Multivariate Behav. Res. 1988. V. 23, iss. 2. P. 231–242. DOI: 10.1207/s15327906mbr2302\_6.
25. Murray G., Carenini G., Ng R. Using the omega index for evaluating abstractive community detection // Proc. of Workshop on Evaluation Metrics and System Comparison for Automatic Summarization, Montréal (Canada), 2012. Assoc. for Comput. Linguistics. P. 10–18.
26. Hanteer O., Rossi L. The meaning of dissimilar: An evaluation of various similarity quantification approaches used to evaluate community detection solutions // Proc. of the IEEE/ACM Intern. conf. on Advances in Social Networks Analysis and Mining, Vancouver (Canada), 2019. P. 513–518. DOI: 10.1145/3341161.3342941.
27. Berlingerio M., Coscia M., Giannotti F. Finding and characterizing communities in multidimensional networks // Intern. conf. on Advances in Social Networks Analysis and Mining (ASONAM). P. 490–494. IEEE Computer Society Washington, DC, USA, 2011. DOI: 10.1107/ASONAM.2011.104.
28. Kim J., Lee J.-G., Lim S. Differential flattening: A novel framework for community detection in multi-layer graphs // ACM Trans. on Intell. Syst. and Technol. (TIST). 2016. V. 8, iss. 2. P. 27:1–27:23. DOI: 10.1145/2898362.
29. De Domenico M., Nicosia V., Arenas A., Latora V. Structural reducibility of multilayer networks // Nature Commun. 2015. V. 6. 6864. DOI: 10.1038/ncomms7864.
30. Bianconi G. Statistical mechanics of multiplex networks: entropy and overlap // Phys. Rev. E. 2013. V. 87, iss. 6. 062806. DOI: 10.1103/PhysRevE.87.062806.
31. Pons P., Latapy M. Computing communities in large networks using random walks. 2006. arXiv: physics/0512106. DOI: 10.48550/arXiv.physics/0512106.

32. Rosvall M., Axelsson D., Bergstrom C. T. Map equation. // Eur. Phys. J. 2009. V. 178. P. 13–23. DOI: 10.1140/epjst/e2010-01179-1.



**Бредихин Сергей Всеволодович** — канд. техн. наук, ведущий научный сотрудник Ин-та вычислительной математики и математической геофизики СО РАН; e-mail:

bredikhin@sscc.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. Лауреат государственной премии по науке и технике 2012 г. Автор и соавтор многих научных статей, докладов 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).



**Щербакова Наталия Григорьевна** — ст. науч. сотр. Ин-та вычислительной математики и математической геофизики СО РАН; e-mail: scherbakova@sscc.ru.

**Наталия 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 электротехники 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. Автор и соавтор более 60 работ и двух монографий.

**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 the co-author of two monographs.

*Дата поступления — 12.08.2025*