



SCIENTIFIC JOURNAL CO-AUTHORSHIP NETWORK MODEL

S. V. Bredikhin, N. G. Scherbakova

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

DOI: 10.24412/2073-0667-2023-3-5-18

EDN: QADKBW

The traditional representation of a complex system (CS) in the form of a network or a graph makes it possible to identify many aspects of the system's behavior, see [2–4]. The network is based on pairwise relationships between network objects. In this case, the nodes of the co-authorship network correspond to the authors, and the edge between the nodes is constructed if there is at least one publication in which both authors participated, see the fundamental works [5, 6]. Such a model is able to express many complex properties of CS, its advantage lies in its simplicity and the ability to use graph theory in the analysis. The model has been studied quite well in terms of the distribution of degrees of network nodes [5], ranking of authors [6], dynamics of evolution [7, 8], identification of communities [9], and prediction of new co-author relations [10, 11].

However, this CS modeling does not reveal all the information that can be extracted directly from the list of articles and their authors, for example, does not reflect the total number of articles prepared by co-authors. To overcome this shortcoming, several methods for constructing networks of co-authorship, extending the traditional approach, were considered, see [6, 12]. But fixing only binary relations between system objects excludes interaction involving groups of network objects. A possible solution to the problem is to generalize the pair interaction to the interaction of an arbitrary set of nodes. In this case, the networks are called hypernetworks. The main mathematical structures used as a CS model include bipartite graphs, hypergraphs, and simplicial complexes [13].

The concept of a hypergraph [20] was proposed in [21] as applied to the analysis of CSs and providing a computationally feasible tool that can be adapted to many analytical situations. In the study of co-authorship, as a rule, the nodes corresponding to the authors form a hyperedge in the case of a joint publication [22, 23].

This paper presents a co-authorship network model that takes into account group relationships that arise between co-authors. The network is modeled using a hypergraph whose vertices correspond to authors and whose edges correspond to scientific articles (SA). The data we analyze is extracted from the electronic archive (<https://www.dia-endojournals.ru/jour/issue/archive>) of the quarterly scientific and practical medical peer-reviewed journal Diabetes Mellitus (ISSN 2072-0378), published since 1998. The journal is indexed in the international abstract and full-text databases. The hypergraph H^{ca} , built on the basis of the selected archive data, has size $m(H^{ca}) = 991$ and order $n(H^{ca}) = 1694$. Since only those SAs that have two or more authors are considered, there are no loops in the constructed hypergraph. The hypergraph H^{ca} is not connected, it consists of 97 components, of which 68 components have one edge, the maximum component includes about 64.75 % vertices (authors) and 67.4 % edges (articles).

The parameters of the hypergraph (and its components) are measured and its topological properties (simple, star, strong star, conformal, Helly) are revealed. The use of the hypergraph language allows to

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

get an idea of the shape of cooperation in the system under consideration. Most of the authors of the co-authored journal articles are indirectly related to each other due to the presence of joint works. The vertex degree distributions (the maximal component) have a long tail, i.e. most authors have a small number of joint SAs. The same can be said about the distribution of edge degrees, indicating that most SAs have a small number of co-authors. All components, except for the maximal, have one or another considered property. Moreover, the properties of conformity and Helly are common to all components, regardless of size and order. The property of simplicity is more characteristic of components with a small number of edges, the same can be said about the probability of being a star. Two components have all the considered properties.

It should be noted that based on the incidence matrix of a hypergraph, a traditional network of co-authorship can be built, in which two authors are connected if they have at least one joint work. The associated multigraph $G(H^{ca})$ is one of the models of such a network in which the weight of an edge between two authors is equal to the number of joint works. This work continues the study and approbation of methods for analyzing networks of co-authorship, see [1].

Key words: complex network, hypergraph, co-authorship, archive of scientific articles, bibliometry.

References

1. BREDIKHIN S. V., LYAPUNOV V. M., SCHERBAKOVA N. G. The hypernetwork of scientific co-authorship. DB RePEc data analysis // Problemy informatiki. 2022. № 4. S. 70–83. DOI: 10.24412/2073-0667-2022-4-70-83.
2. NEWMAN M. E. J. Networks: An introduction. Oxford Univ. Press, 2010. ISBN: 9780199206650.
3. ESTRADA E. The structure of complex network: Theory and applications. Oxford Univ. Press, 2011. ISBN: 9780199591756.
4. LATORA V., NICOSIA V., RUSSO G. Complex networks: Principles, methods and applications. Cambridge Univ. Press, 2017. ISBN: 9781316216002.
5. NEWMAN, M. E. J. Scientific collaboration networks. I. Network construction and fundamental results // Phys. Rev. 2001. E 64, 016131. DOI: 10.1103/PhysRevE.64.016131.
6. NEWMAN, M. E. J. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality // Phys. Rev. 2001. E 64, 016132. DOI: 10.1103/PhysRevE.64.016132.
7. BARABASI A. L., JEONG H., NEDA Z., RAVASZ E., SCHUBERT A., VICSEK T. Evolution of the social network of scientific collaborations // Physica A. 2002. V. 311. P. 590–614. DOI: 10.1016/S0378-4371(02)00736-7.
8. UDDIN S., HOSSAIN L., ABBASI A., RASMUSSEN K. Trend and efficiency analysis of co-authorship network // Scientometrics. 2012. V. 90, No 2. P. 687–699. DOI: 10.1007/s11192-011-0511-x.
9. SAVIĆ M., IVANOVIC M., RADOVANOVIC M., OGNJANOVIC Z., PEJOVIĆ A. Exploratory analysis of communities in co-authorship networks: A case study // Intern. Conf. on ICT Innovations. Springer, 2019. P. 55–64. ISBN 978-3-319-91194-6.
10. CHO H., YU Y. Link prediction for interdisciplinary collaboration via co-authorship network // Soc. Network Analysis and Mining. 2018. 8, 25. DOI: 10.1007/s13278-018-0501-6.
11. CHUAN P.M., SON L.H., ALI M. ET AL. Link prediction in co-authorship networks based on hybrid content similarity metric // Appl. Intell. 2018. V. 48. P. 2470–2486. DOI: 10.1007/s10489-017-1086-x.
12. 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. DOI: 10.1007/s11192-006-0113-1.
13. SCHERBAKOVA N. G. Modeling group interactions of complex systems. Review // Problemy informatiki. 2022. № 3. S. 24–45. DOI: 10.24412/2073-0667-2022-3-24-45.

14. WILSON T. P. Relational networks: An extension of sociometric concepts // Soc. Networks. 1982. V. 4, iss. 2. P. 105–116. DOI: 10.1016/0378-8733(82)90028-4.
15. BORGATTI S. P., EVERETT M. G. Network analysis of 2-mode data // Soc. networks. 1997. V. 19. P. 243–269. DOI: 10.1016/S0378-8733(96)00301-2.
16. FAUST K. Centrality in affiliation networks // Soc. Networks. 1997. V. 19. P. 157–191. DOI: 10.1016/S0378-8733(96)00300-0.
17. WASSERMAN S., FAUST K. Social network analysis. Cambridge Univ. Press, 1984. ISBN 9780511815478.
18. BENKO G., FLAMM C., STADLER P. F. Generic properties of chemical networks: Artificial chemistry based on graph rewriting // Lect. Notes Comput. Sci. 2003. V. 2801, 10. DOI: 10.1007/978-3-540-39432-7_2.
19. BANERJEE S., JENAMANI M., PRATHAR D. K. Properties of a projected network of a bipartite network // arXiv:1707.00912v1. DOI: 10.48550/arXiv.1707.00912.
20. BERGE C. Graphs and hypergraphs. Amsterdam: North-Holland, 1976. ISBN 10.0444103996.
21. SEIDMAN S. Structures induced by collections of subsets: A hypergraph approach // Math. Soc. Sci. 1981. V. 1, iss. 4. P. 381–396. DOI: 10.1016/0165-4896(81)90016-0.
22. ESTRADA E., RODRÍGUEZ-VELÁZQUEZ J. A. Complex networks as hypergraphs // arXiv: physics/0505137, 2005. DOI: 10.1016/j.physa.2005.12.002.
23. HAN Y., ZHOU B., PEI J., JIA Y. Understanding importance of collaborations in coauthorship networks: A supportiveness analysis approach // Proc. 2009 SIAM Internat. Conf. on Data Mining. 2009. P. 1112–1123. DOI: 10.1137/1.9781611972795.95.
24. LUNG R. I., GASCO N., SUCIU M. A. A hypergraph model for representing scientific output // Scientometrics. 2018. V. 117. P. 1361–1379. DOI: 10.1007/s11192-018-2908-2.
25. ATKIN R. H. From cohomology in physics to q-connectivity in social science // Intern. J. Man-Machine Studies. 1972. V. 4, iss. 2. P. 139–167. DOI: 10.1016/S0020-7373(72)80029-4.
26. ATKIN R. H. An algebra for patterns on a complex, I // Intern. J. Man-Machine Studies. 1974. V. 6, iss. 3. P. 285–307. DOI: 10.1016/S0020-7373(74)80024-6.
27. ATKIN R. H. An algebra for patterns on a complex, II // Intern. J. Man-Machine Stud. 1976. V. 8, iss. 5. P. 483–498. DOI: 10.1016/S0020-7373(76)80015-6.
28. PATANIA A., PETRI G., VACCARINO F. The shape of collaborations // EPJ Data Science. 2017. V. 6, 18. DOI 10.1140/epjds/s13688-017-0114-8.
29. ZHOU D., ORSHANSKIY S. A., ZHA H., GILES C. L. Co-ranking authors and documents in a heterogeneous network // 7th IEEE International Conference on Data Mining (ICDM 2007), Omaha (USA), 2007. P. 739–744. DOI: 10.1109/ICDM.2007.57.
30. VOLOSHIN V. I. Introduction to graph and hypergraph theory. NY: Nova Science Publ., 2009. ISBN: 978-1-60692372-6.
31. BRETT A. Hypergraph theory. Springer, 2013. ISBN 978-3-319-00079-4.
32. CORMEN T. H., LEISERSON C., RIVEST R. L. Algoritmy: postroenie i analiz. M.: MTSNMO, 2002. ISBN: 978-5-907114-11-1.



МОДЕЛЬ СЕТИ СОАВТОРСТВА НАУЧНОГО ЖУРНАЛА

С. В. Бредихин, Н. Г. Щербакова

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

УДК 519.177

DOI: 10.24412/2073-0667-2023-3-5-18

EDN: QADKBW

Традиционный подход к изучению сети научного соавторства основан на бинарных отношениях, возникающих между авторами, создавшими по крайней мере одну совместную публикацию. В работе представлена модель сети соавторства, учитывая гиперграфа, вершины которого соответствуют авторам, а ребра – публикациям. Представлен метод построения гиперграфа сети соавторства, основанный на данных, извлеченных из архива журнала. Измерены параметры гиперграфа и выявлены его топологические свойства. Данная работа продолжает изучение и апробацию методов анализа сетей соавторства (см. [1]).

Ключевые слова: комплексная сеть, гиперграф, соавторство, архив научных статей, библиометрия.

Введение. Представление комплексной системы (КС) в виде сети позволяет выявить многие аспекты поведения системы [2–4]. Сеть базируется на парных отношениях между сетевыми объектами. В этом случае узлам сети соавторства соответствуют авторы, а ребро между узлами строится в случае наличия хотя бы одной публикации, в которой участвовали оба автора (см. основополагающие работы [5, 6]). Такая модель способна выразить многие сложные структуры и динамические свойства представленной системы, ее преимущество заключается в простоте и возможности использовать при анализе теорию графов. Модель достаточно хорошо изучена с точки зрения распределения степеней узлов сети [5], ранжирования авторов [6], динамики развития [7, 8], выявления сообществ [9], прогнозирования новых соавторских связей [10, 11]. Однако структура сети не раскрывает всей информации, которую мы можем извлечь непосредственно из списка статей и их авторов, например, не отражает общее число статей, подготовленных соавторами.

Для преодоления этого недостатка было рассмотрено несколько методов построения сетей соавторства, расширяющих традиционный подход. Наиболее простой – построение взвешенной сети путем присвоения ребру веса, равного числу статей, написанных двумя исследователями совместно, это так называемый *полный вес*. *Частичный вес* устанавливается с учетом числа авторов каждой совместной статьи [6]. Рассматривались также ориентированные сети соавторства, например учитывающие порядок авторов в заголовке статьи [12]. Однако моделирование комплексной сети графом не обеспечивает полного описания исследуемой системы. Фиксация исключительно бинарных отношений между объектами системы исключает взаимодействие с участием групп сетевых объектов. В случае

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

соавторства мы имеем дело с группой соавторов одной публикации, взаимодействующими одновременно. Кроме того, если несколько авторов образуют клику в графе соавторства, это не значит, что у них есть совместная публикация.

Возможным решением проблемы является обобщение парного взаимодействия до взаимодействия произвольного множества узлов. В этом случае сети принято называть *гиперсетями*. К основным математическим структурам, используемым в качестве модели КС, можно отнести двудольные графы, гиперграфы и симплициальные комплексы [13].

Моделирование КС двудольным графом, предложенное в работе [14], широко принято в социальных сетях [15–17]. В этом случае имеются два непересекающихся множества узлов, одно соответствует публикациям, другое – авторам. Такая неоднородность узлов осложняет исследование центральности, связности и кластеризации [15, 18]. Следует заметить, что большинство методов анализа двудольных сетей основано на использовании проекций [19].

Концепция гиперграфа [20] в применении к анализу КС предложена в работе [21] как предоставляющая вычислительно выполнимый инструмент, который можно адаптировать ко многим аналитическим ситуациям. При исследовании соавторства, как правило, узлы, соответствующие авторам, объединяются гиперребром в случае наличия совместной публикации [22, 23]. Иной подход применен в работе [24], при котором узлы соответствуют публикациям, узлы объединяются гиперребром, соответствующим автору, принявшему участие в публикациях.

В цикле работ [25–27] социальные отношения представляются n -мерными многогранниками – симплексами. Разница с подходом, приведенным в работе [21], заключается в том, что гиперребро, состоящее из $p + 1$ вершины, представляется геометрической структурой, p -симплексом $\sigma = \langle v_0, v_1, \dots, v_p \rangle$, а КС – множеством симплексов, замкнутым относительно включения всех подсимплексов всех симплексов, т. е. если σ входит в комплекс, то все $\sigma' \subset \sigma$ входят в комплекс. В случае соавторства можно считать, что каждая публикация – это симплекс, определяемый множеством соавторов. В отличие от гиперграфа симплекс образует максимальная группа авторов, не являющаяся подгруппой авторов других публикаций [28]. Топологическое описание данных о сотрудничестве позволяет выявить связь между формами отношений между сетевыми сообществами и структурой ассоциированного топологического объекта.

Выбор метода моделирования множественных отношений зависит от решаемой задачи. Так, в работе [29] проблема оценки научных публикаций и авторов решается путем исследования трех сетей: социальной сети, объединяющей авторов посредством соавторства, сети цитирования, объединяющей публикации, и двудольной сети соавторства, объединяющей авторов с публикациями. Наличие топологических пустот рассматривается в работе [28] как важная информация при выявлении форм сотрудничества, и этот метод лучше всего подходит для симплициальных симплексов. В то же время если при исследовании системы мы задаемся вопросом, работала ли именно данная группа авторов над публикацией и какой именно автор принимал участие в большем числе проектов, то следует опираться на гиперграф. Понятие гиперграфа включает как частные случаи широкий спектр других математических структур, которые подходят для изучения сложных сетей [22].

В данной работе система соавторства представлена в виде гиперсети, в которой узлы соответствуют авторам, а ребра – публикациям. Соавторы одной публикации объединяются в ребро. Исследуются характеристики сети, позволяющие расширить ин-

формацию о системе соавторства в сравнении с методикой представления на основе графа.

1. Метод исследования 1.1. *Гиперграф. Определения и утверждения.* Приведем формальные сведения, необходимые далее для исследования соавторства, почерпнутые из монографий [20, 30, 31].

Гиперграф $H = (V, E)$ на конечном множестве $V = \{v_1, v_2, \dots, v_n\}$ определяется семейством $E = (E_1, E_2, \dots, E_m)$ подмножеств множества V , таких что: а) $E_i \neq \emptyset$, $i = 1, 2, \dots, m$; б) $\bigcup_{i=1}^m E_i = V$. Элементы v_1, v_2, \dots, v_n называются *вершинами* H , а E_1, E_2, \dots, E_m – *ребрами* или *гиперребрами*. Число вершин H называется *порядком* гиперграфа и обозначается n , а число ребер называется *размером* и обозначается m . Ранг гиперграфа $r(H)$ определяется как $r(H) = \max_j |E_j|$. Вершины $v_i, v_k \in V$ называются *смежными*, если существует ребро $E_j \in E$, такое что $v_i \in E_j$ и $v_k \in E_j$. Если $v_i \in E_j$, то ребро и вершина *инцидентны*. Обозначим $E(v_i)$ множество всех ребер, содержащих вершину v_i , назовем это множество *звездой* с центром v_i :

$$E(v_i) = \{E_j \in E : v_i \in E_j\}.$$

Число $|E(v_i)|$ называется *степенью вершины* v_i , обозначается $\deg(v_i)$. Число $|E_i|$ называют *степенью (размером, мощностью) ребра* E_i .

Матрицей инцидентности гиперграфа H называется $(0, 1)$ -матрица $C(H) = (c_{ij})$ с n строками и m столбцами, определенная как:

$$c_{ij} = \begin{cases} 0, & \text{если } v_i \notin E_j ; \\ 1, & \text{если } v_i \in E_j. \end{cases}$$

В терминах $C(H)$ степенью вершины является сумма элементов соответствующей строки матрицы инцидентности. Матрицей смежности гиперграфа $A(H) = (a_{ij})$ называется квадратная матрица, элемент a_{ij} – это число ребер, содержащих одновременно вершины v_i и v_j , диагональные элементы матрицы A равны нулю:

$$(\forall v_i \in V) a_{ii} = 0; \quad (\forall v_i, v_j \in V, i \neq j) a_{ij} = |E_k \in E : \{v_i, v_j\} \subseteq E_k|.$$

Матрица $A(H)$ может быть получена из матрицы $C(H)$ следующим образом:

$$A(H) = C \cdot C^\top - D_v,$$

где C^\top – транспонированная матрица инцидентности, D_v – матрица, диагональные элементы которой равны степеням соответствующих вершин гиперграфа: $d_{ii} = \deg(v_i)$. Матрицу $A(H)$ можно рассматривать как матрицу смежности мультиграфа $G(H)$ (т. е. взвешенного графа), называемого *ассоциированным графом* гиперграфа $H = (V, E)$. Заметим, что в монографии [30] матрица смежности – это $(0, 1)$ -матрица:

$$a_{ij, i \neq j} = 1, \text{ если } E(v_i) \cap E(v_j) \neq \emptyset.$$

Двойственным гиперграфу $H = (V, E)$ называется гиперграф $H^* = (V^*, E^*)$, такой что его вершины $V^* = e_1, e_2, \dots, e_m$ соответствуют ребрам в H , а ребра H^* соответствуют вершинам H с отношением инцидентности, связывающим каждую вершину с ребрами H , в

которые вершина входит: $E_i^* = \{e_j : v_i \in E_j \text{ в } H\}$. Матрица инцидентности C^* гиперграфа H^* может быть вычислена как $C^* = C^\top$, таким образом $(H^*)^* = H$. Двойственный гиперграфу H без повторяющихся ребер гиперграф H^* может иметь повторяющиеся ребра, а также оказаться обычным графом (примеры см. [30, 31]).

Реберным графом $L(H)$ гиперграфа H называется граф, множество вершин которого совпадает с множеством ребер H , а две вершины являются смежными тогда и только тогда, когда соответствующие ребра H пересекаются:

$$L(H) = (V', E'), \text{ где } V' = E; \{E_i, E_j\} \in E' \Leftrightarrow E_i \cap E_j \neq \emptyset.$$

Здесь и далее для простоты E_j используется как обозначение ребра и как множество вершин, входящих в ребро.

Для гиперграфа $H = (V, E)$ *2-секция* [30] – это граф, обозначаемый как $[H]_2$, множество вершин которого совпадает с множеством вершин гиперграфа H , а две различные вершины являются смежными тогда и только тогда, когда в H существует ребро, которому принадлежат обе вершины:

$$[H]_2 = (V, E_2), \text{ где } \{v_i, v_j\} \in E_2 \Leftrightarrow E(v_i) \cap E(v_j) \neq \emptyset.$$

Для любого гиперграфа H выполняются соотношения [30]:

$$[H]_2 = L(H^*); \quad L(H) = [H^*]_2.$$

Матрицу смежности гиперграфа в определении [30] можно рассматривать как матрицу смежности графа $[H]_2$. В монографии [31] приведено расширение понятия 2-секции гиперграфа: $H = (V, E)$ – это мультиграф с помеченными ребрами, в котором множество вершин совпадает с множеством вершин гиперграфа H , а вершины v_i и v_j связаны ребром, помеченым e , если $v_i, v_j \subseteq e \in E$.

В гиперграфе H чередующаяся последовательность $p = v_0 E_0 v_1 E_1 v_2 \dots v_{l-1} E_{l-1} v_l$ различных вершин $v_0, v_1, v_2, \dots, v_{l-1}, v_l$ и различных ребер $E_0, E_1, E_2, \dots, E_{l-1}$, удовлетворяющая условию $(v_i, v_{i+1}) \in E_i$, $i = 0, 1, \dots, l - 1$, называется *путем*, или (v_0, v_l) -*путем*, связывающим вершины v_0, v_l , и называется *циклом*, если $v_0 = v_l$. Величина l называется *длиной пути*. *Дистанция* между вершинами – это длина кратчайшего пути между ними.

Гиперграф H называется *связным*, если для каждой пары вершин существует связывающий их путь. Если гиперграф несвязный, то существуют две или более *связные компоненты*, каждая из которых представляет собой гиперграф. Изолированная вершина, а также вершина, инцидентная только одной петле, также являются связными компонентами.

Подсемейство E_1, \dots, E_k ребер гиперграфа $H = (V, E)$ является *пересекающимся семейством*, если каждая пара ребер этого подсемейства имеет непустое пересечение:

$$\forall i, j \in [k] \quad E_i \cap E_j \neq \emptyset.$$

Гиперграф называется *пересекающимся семейством*, если все его ребра попарно пересекаются. Гиперграф $H = (V, E)$ называется *звездой*, если существует вершина, принадлежащая всем его ребрам:

$$\exists v_i \forall E_{j \in [n]} \quad v_i \in E_j,$$

т. е. гиперграф является пересекающимся семейством с непустым пересечением всех ребер.

Гиперграф звезда является *строгой звездой*, если только вершины, общие для любой пары пересекающихся ребер, составляют пересечение всех ребер.

Гиперграф обладает *свойством Хелли* (*является Хелли*), если каждое его пересекающееся семейство ребер имеет непустое пересечение, т. е. для каждого семейства E_1, \dots, E_k из k ребер выполняется условие

$$\forall i, j \in [k] \quad E_i \cap E_j \neq \emptyset \Rightarrow E_1 \cap E_2 \cap \dots \cap E_k \neq \emptyset.$$

Кликой [32] в неориентированном графе называется подмножество вершин, каждые две из которых соединены ребром графа. Клика *максимальная (по включению)*, если ее нельзя расширить добавлением в нее вершин. Гиперграф H , в котором ребра являются максимальными кликами $[H]_2$, называется *конформным*. Если гиперграф не конформный, то при переходе от гиперграфа к графу парных отношений не все структуры графа возникают на основании структуры гиперграфа.

Приведем два утверждения из монографии [30], позволяющие выявлять конформность без перехода к 2-секции.

Утверждение 1 (теорема Гилмора). Гиперграф $H = (V, E)$ конформный тогда и только тогда, когда для любых попарно пересекающихся трех ребер E_1, E_2, E_3 существует ребро E_4 , такое что выполняется $(E_1 \cap E_2) \cup (E_1 \cap E_3) \cup (E_2 \cap E_3) \subseteq E_4$.

Утверждение 2. Гиперграф H конформный тогда и только тогда, когда двойственный граф H^* является Хелли.

1.2. *Моделирование сети соавторства.* Исходные данные о соавторстве содержатся в научных статьях (НС), хранящихся в архиве со свободным доступом. Обозначим $P = \{p_1, p_2, \dots, p_m\}$ множество НС, $S = \{s_1, s_2, \dots, s_n\}$ – множество их авторов. Предполагаем, что P содержит только те НС, которые имеют двух и более авторов, т. е. в конструируемом гиперграфе отсутствуют петли. Далее не будем различать термины гиперсеть и гиперграф. Построим гиперграф $H = (V, E)$, такой что множество S отображается на множество вершин V , а множество P – на множество ребер E , причем все авторы одной публикации образуют ребро, соответствующее публикации: если НС p_i подготовлена именно авторами s_1, s_2, \dots, s_k , то $E_i = \{v_1, v_2, \dots, v_k\}$ является ребром, $E_i \in E$ (см. [22]).

Исследуем параметры и свойства гиперграфа H и/или его компонент. Для максимальной компоненты рассмотрен ряд свойств двойственного гиперграфа H^* .

Параметры:

Размер. При таком подходе гиперграф будет иметь размер $m(H) = |P|$ и порядок $n(H) = |S|$.

Ранг. Ранг указывает на максимальное число соавторов у НС архива.

Средняя степень ребра. Это среднее число соавторов у НС.

Распределение степеней вершин. Это распределение числа совместных НС, в которых участвовал соответствующий автор.

Свойства:

Связность. Число, размер и состав компонент. Размеры максимальной компоненты свидетельствуют об опосредованной связи между авторами, а также между НС на основе соавторства.

Простота. Если гиперграф не является простым, то наличие кратных ребер указывает на то, что имеются конкретные группы авторов, подготовившие совместно несколько

публикаций, а наличие вложенных ребер – что в группе авторов имеются подгруппы, подготовившие работы в более узком составе.

Петли. В гиперграфе H нет петель, так как рассматриваются только НС, имеющие более одного автора, но они могут быть у двойственного гиперграфа.

Звезда. Все публикации имеют хотя бы одного общего автора.

Строгая звезда. Все публикации имеют одну общую группу соавторов.

Конформность. В этом случае тесные группы авторов (клики графа парных отношений соавторства, обычно попадающих в одно сообщество при кластеризации) входят в число соавторов хотя бы одной НС.

Хелли. Для любого множества НС, такого что каждая пара НС имеет общих авторов, выполняется: все НС множества имеют хотя бы одного общего автора (входят в звезду).

2. Исходные данные. Данные, которые мы анализируем, извлечены из электронного архива ежеквартального научно-практического медицинского рецензируемого журнала «Сахарный диабет» (ISSN 2072-0378), издаваемого с 1998 г. Журнал индексируется в международных реферативных и полнотекстовых базах данных (Web of Science, Scopus, Google Scholar, включен в Перечень ВАК). Архив журнала (<https://www.dia-endojournals.ru/jour/issue/archive>) содержит XML массив научных статей, опубликованных за период с 2007 (том 10, № 1) по 2023 годы (том 26, № 1), и находится в свободном доступе.

Каждой НС соответствует размеченный текстовый файл, состоящий из записей вида (фрагмент):

```
<article-title>Название статьи</article-title>
<contrib contrib-type="author-n">Автор-n </contrib>
<aff id="aff-m">Аффилиация-m </aff>
<abstract>Аннотация</abstract>
<kwd-group>Ключевые слова</kwd-group>
<body>Текст статьи</body>
<ref-list>Список литературы</ref-list>
```

Здесь *n* – порядковый номер автора в пристатейном списке авторов НС, *m* – номер его аффилиации.

В январе 2023 г. из архива журнала было извлечено 1067 XML файлов, соответствующих НС, опубликованным за период 2007–2023 гг. Предварительная обработка «сырых» данных состояла в следующем: во-первых, были удалены НС, имеющие одного автора, во-вторых, введен единообразный порядок следования фамилии, имени и отчества автора. В результате было отснято менее двух процентов НС.

3. Результат. Гиперграф H^{ca} , построенный на основе отобранных данных архива, имеет размер $m(H^{ca}) = 991$ и порядок $n(H^{ca}) = 1694$. H^{ca} не является связным, он состоит из 97 компонент, из них 68 имеют одно ребро, максимальная компонента включает около 64,75 % вершин (авторов) и 67,4 % ребер (НС).

Измерим параметры и определим свойства компонент с числом ребер не менее трех, рассмотренных как отдельные гиперграфы. При этом максимальную компоненту обозначим H^{ca} , для нее приведем параметры и свойства двойственного гиперграфа H^* , остальные компоненты пронумерованы. Значения параметров и свойства гиперграфов приведены в таблице. Распределение степеней вершин H^{ca} представлено на рис. 1 (ось *x* – степени вершин, ось *y* – плотность; шкалы логарифмические). Из графика распределения можно заключить, что большинство авторов имеют небольшое число совместных НС. Распре-

Таблица 1

Параметры и свойства компонент

Компонента имя/номер	Параметры				Свойства				
	Размер $m(\cdot)$	Порядок $n(\cdot)$	Ранг $r(\cdot)$	$avg_deg(\cdot)$	Простой	Звезда	Строгая звезда	Конформный	Хелли
H^{ca}	668	1097	38	4,8	—	—	—	—	—
H^*	1097	668	157	2,9	—	—	—	—	—
3	19	60	8	5,7	—	—	—	+	+
4	9	33	10	5,4	—	—	—	+	+
5	5	20	9	6,2	—	—	—	+	+
6	6	15	7	4,3	—	—	—	+	+
7	5	15	8	4,4	—	+	—	+	+
8	3	15	8	5,6	+	—	—	+	+
9	3	13	6	5,6	+	+	—	+	+
10	3	13	6	5,3	+	—	—	+	+
11	3	12	6	4,7	+	+	+	+	+
12	4	9	6	3,5	+	+	—	+	+
13	4	9	4	3,3	—	+	—	+	+
14	3	8	5	4	—	+	—	+	+
15	3	8	4	3,7	+	+	—	+	+
16	4	7	4	2,8	—	+	—	+	+
17	3	7	4	3	+	+	+	+	+
18	3	7	5	4,3	—	+	—	+	+
19	4	5	4	3	—	+	—	+	+
20	3	4	2	2	+	—	—	+	+

Примечание: $avg_deg(\cdot)$ – среднее значение степени ребра; знак “+” означает, что гиперграф обладает указанным свойством, знак “–”, что не обладает

деление степеней вершин H^* также является распределением с длинным хвостом, что свидетельствует о том, что большинство НС имеет небольшое число соавторов.

Выводы. Все компоненты не имеют петель, так как рассматривались НС, имеющие более одного автора, однако гиперграф H^* , двойственный H^{ca} , имеет петли (то же для двойственного гиперграфа компоненты 3, следующей по размерам за максимальной), а значит существуют авторы, которые имеют только одну НС, подготовленную в соавторстве. Как видно из таблицы, гиперграф H^{ca} , соответствующий максимальной компоненте, и двойственный гиперграф H^* не обладают ни одним из перечисленных в таблице свойств, в то время как гиперграф, соответствующий компоненте 3, обладает свойствами конформности и Хелли. Две компоненты (11 и 17) обладают всеми свойствами. Они являются строгими звездами с тремя ребрами, ребра имеют в пересечении точно одну вершину, т. е. состоят из НС, объединенных одним автором. Компонента 17 представлена на рис. 2. Более половины компонент являются звездами, т. е. все ребра (НС) имеют в пересечении хотя бы одну вершину. Все компоненты, кроме максимальной, являются конформными, это означает, что при переходе к ассоциированному графу не будет клик, которые ошибочно могут считаться тесными сообществами.

Заключение. В статье приведены результаты анализа топологической структуры данных о соавторстве, извлеченных из архива, находящегося в свободном доступе. В качестве формализма для кодирования комплексной системы соавторства, отражающего множественные отношения между авторами, используется концепция гиперграфа. Применение

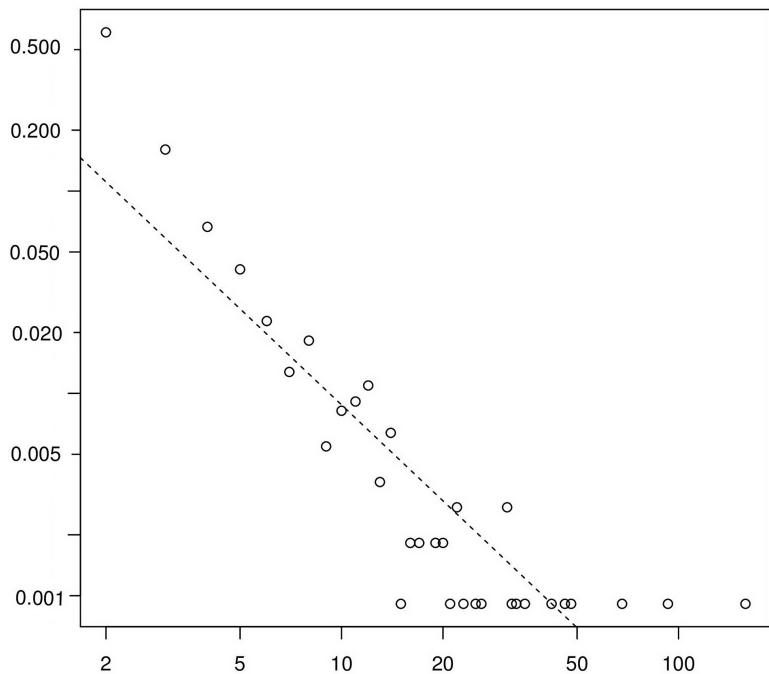
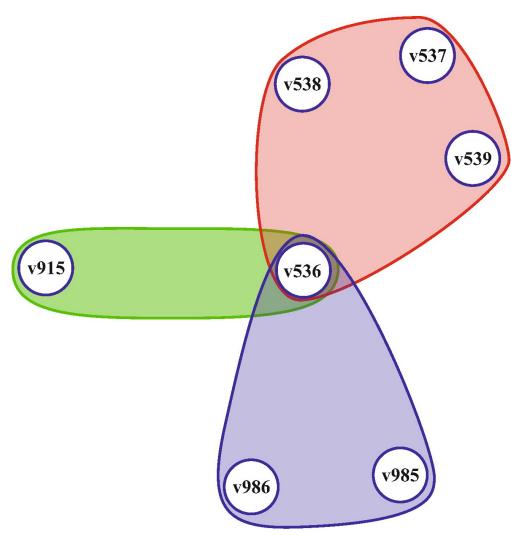
Рис. 1. Распределение степеней вершин гиперграфа H^{ca} 

Рис. 2. Компонента 17

языка гиперграфов позволяет получить представление о шаблонах сотрудничества в рассматриваемой системе. Большинство авторов рассматриваемой системы соавторства опосредованно связаны между собой благодаря наличию совместных работ. Распределения степеней вершин максимальной компоненты имеет длинный хвост, т. е. большинство авторов имеют небольшое число совместных НС. То же можно утверждать о распределении степеней ребер, следовательно, большинство НС имеют небольшое число соавторов. Все компоненты, кроме максимальной, обладают тем или иным заявленным свойством. Причем свойство конформности и Хелли присущи всем компонентам вне зависимости от размера и порядка. Свойство простоты более характерно для компонент с небольшим числом ребер, это относится и к вероятности быть звездой. Две компоненты обладают всеми рассматриваемыми свойствами.

Следует заметить, что на основе матрицы инцидентности гиперграфа может быть построена традиционная сеть соавторства, в которой два автора связаны, если у них имеется хотя бы одна совместная работа. Ассоциированный мультиграф граф $G(H^{ca})$ является одной из моделей такой сети, в которой вес ребра между двумя авторами равен числу совместных работ.

Благодарности. Авторы благодарят А. В. Феофанова за помощь в парсинге исходных данных.

Список литературы

1. БРЕДИХИН С. В., ЛЯПУНОВ В. М., ЩЕРБАКОВА Н. Г. Гиперсеть научного соавторства. Анализ данных БД Repes // Пробл. информ. 2022. № 4. С. 70–83. DOI: 10.24412/2073-0667-2022-4-70-83.
2. NEWMAN M. E. J. Networks: An introduction. Oxford Univ. Press, 2010. ISBN: 9780199206650.

3. ESTRADA E. The structure of complex network: Theory and applications. Oxford Univ. Press, 2011. ISBN: 9780199591756.
4. LATORA V., NICOSIA V., RUSSO G. Complex networks: Principles, methods and applications. Cambridge Univ. Press, 2017. ISBN: 9781316216002.
5. NEWMAN, M. E. J. Scientific collaboration networks. I. Network construction and fundamental results // Phys. Rev. 2001. E 64, 016131. DOI: 10.1103/PhysRevE.64.016131.
6. NEWMAN, M. E. J. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality // Phys. Rev. 2001. E 64, 016132. DOI: 10.1103/PhysRevE.64.016132.
7. BARABÁSI A. L., JEONG H., NÉDA Z., RAVASZ E., SCHUBERT A., VICSEK T. Evolution of the social network of scientific collaborations // Physica A. 2002. V. 311. P. 590–614. DOI: 10.1016/S0378-4371(02)00736-7.
8. UDDIN S., HOSSAIN L., ABBASI A., RASMUSSEN K. Trend and efficiency analysis of co-authorship network // Scientometrics. 2012. V. 90, No 2. P. 687–699. DOI: 10.1007/s11192-011-0511-x.
9. SAVIĆ M., IVANOVIC M., RADOVANOVIC M., OGNJANOVIC Z., PEJOVIĆ A. Exploratory analysis of communities in co-authorship networks: A case study // Intern. Conf. on ICT Innovations. Springer, 2019. P. 55–64. ISBN 978-3-319-91194-6.
10. CHO H., YU Y. Link prediction for interdisciplinary collaboration via co-authorship network // Soc. Network Analysis and Mining. 2018. 8, 25. DOI: 10.1007/s13278-018-0501-6.
11. CHUAN P.M., SON L.H., ALI M. ET AL. Link prediction in co-authorship networks based on hybrid content similarity metric // Appl. Intell. 2018. V. 48. P. 2470–2486. DOI: 10.1007/s10489-017-1086-x.
12. 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. DOI: 10.1007/s11192-006-0113-1.
13. ЩЕРБАКОВА Н. Г. Моделирование групповых взаимодействий комплексных систем. Обзор // Пробл. информ. 2022. № 3. С. 24–45. DOI: 10.24412/2073-0667-2022-3-24-45.
14. WILSON T. P. Relational networks: An extension of sociometric concepts // Soc. Networks. 1982. V. 4, iss. 2. P. 105–116. DOI: 10.1016/0378-8733(82)90028-4.
15. BORGATTI S. P., EVERETT M. G. Network analysis of 2-mode data // Soc. networks. 1997. V. 19. P. 243–269. DOI: 10.1016/S0378-8733(96)00301-2.
16. FAUST K. Centrality in affiliation networks // Soc. Networks. 1997. V. 19. P. 157–191. DOI: 10.1016/S0378-8733(96)00300-0.
17. WASSERMAN S., FAUST K. Social network analysis. Cambridge Univ. Press, 1984. ISBN 9780511815478.
18. BENKO G., FLAMM C., STADLER P. F. Generic properties of chemical networks: Artificial chemistry based on graph rewriting // Lect. Notes Comput. Sci. 2003. V. 2801, 10. DOI: 10.1007/978-3-540-39432-7_2.
19. BANERJEE S., JENAMANI M., PRATHAR D. K. Properties of a projected network of a bipartite network // arXiv:1707.00912v1. DOI: 10.48550/arXiv.1707.00912.
20. BERGE C. Graphs and hypergraphs. Amsterdam: North-Holland, 1976. ISBN 10.0444103996.
21. SEIDMAN S. Structures induced by collections of subsets: A hypergraph approach // Math. Soc. Sci. 1981. V. 1, iss. 4. P. 381–396. DOI: 10.1016/0165-4896(81)90016-0.
22. ESTRADA E., RODRÍGUEZ-VELÁZQUEZ J. A. Complex networks as hypergraphs // arXiv: physics/0505137, 2005. DOI: 10.1016/j.physa.2005.12.002.
23. HAN Y., ZHOU B., PEI J., JIA Y. Understanding importance of collaborations in coauthorship networks: A supportiveness analysis approach // Proc. 2009 SIAM Internat. Conf. on Data Mining. 2009. P. 1112–1123. DOI: 10.1137/1.9781611972795.95.
24. LUNG R. I., GASCO N., SUCIU M. A. A hypergraph model for representing scientific output // Scientometrics. 2018. V. 117. P. 1361–1379. DOI: 10.1007/s11192-018-2908-2.

25. ATKIN R. H. From cohomology in physics to q-connectivity in social science // Intern. J. Man-Machine Studies. 1972. V. 4, iss. 2. P. 139–167. DOI: 10.1016/S0020-7373(72)80029-4.
26. ATKIN R. H. An algebra for patterns on a complex, I // Intern. J. Man-Machine Studies. 1974. V. 6, iss. 3. P. 285–307. DOI: 10.1016/S0020-7373(74)80024-6.
27. ATKIN R. H. An algebra for patterns on a complex, II // Intern. J. Man-Machine Stud. 1976. V. 8, iss. 5. P. 483–498. DOI: 10.1016/S0020-7373(76)80015-6.
28. PATANIA A., PETRI G., VACCARINO F. The shape of collaborations // EPJ Data Science. 2017. V. 6, 18. DOI 10.1140/epjds/s13688-017-0114-8.
29. ZHOU D., ORSHANSKIY S. A., ZHA H., GILES C. L. Co-ranking authors and documents in a heterogeneous network // 7th IEEE International Conference on Data Mining (ICDM 2007), Omaha (USA), 2007. P. 739–744. DOI: 10.1109/ICDM.2007.57.
30. VOLOSHIN V. I. Introduction to graph and hypergraph theory. NY: Nova Science Publ., 2009. ISBN: 978-1-60692372-6.
31. BRETTO A. Hypergraph theory. Springer, 2013. ISBN 978-3-319-00079-4.
32. КОРМЕН Т., ЛЕЙЗЕРСОН Ч., РИВЕСТ Т. Алгоритмы: построение и анализ. М.: МЦНМО. 2002. ISBN: 978-5-907114-11-1.



**Бредихин Сергей Все-
володович** — канд. техн. на-
ук, зав. лабораторией Ин-та
вычислительной математики и
математической геофизики СО
РАН; e-mail: bred@nsc.ru;

Сергей Бредихин окон-
чил механико-математический факультет Но-
восибирского государственного университета в
1968 г. С 1968 г. — сотрудник Института ав-
томатики и электрометрии СО РАН. Кандидат
технических наук с 1983 г. С 1988 г. — заведую-
щий Лабораторией прикладных систем Инсти-
тута вычислительной математики и математи-
ческой геофизики СО РАН. Являлся техниче-
ским директором проекта «Сеть Интернет Но-
восибирского Научного Центра». Лауреат госу-
дарственной премии по науке и технике 2012 г.
В сфере его научных интересов — измерение
и анализ сетей распределенных информацион-
ных структур. Автор и соавтор более 110 работ
и двух монографий: «Методы библиометрии и
рынок электронной научной периодики», «Ана-
лиз 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 rynok elektronnoj nauchnoj periodiki”, “Ansliz tsitirovaniya v bibliometrii”.



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

Наталья Щербакова
окончила Новосибирский го-
сударственный университет по
специальности «Математическая лингвистика»
в 1967 г. С 1967 г. работала в Институте ма-
тематики СО РАН, затем в Институте авто-
матики и электрометрии СО РАН в области
создания программного обеспечения систем пе-
редачи данных. С 2000 г. — сотрудник Институ-
та вычислительной математики и математиче-
ской 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. Ав-

тор 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 “Analiz 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 — 28.08.2023