

ВЛИЯНИЕ КОЛЛАБОРАЦИЙ НА ХАРАКТЕРИСТИКИ ГРАФА СОАВТОРСТВА В ОБЛАСТИ ФИЗИКИ В РОССИИ

Е. Е. Васильева¹, О. В. Иванов¹, А. М. Коваленко¹, А. В. Колобов¹,
В. В. Королева¹, А. В. Леонидов^{1,2}

В представленной работе рассматриваются топологические свойства графа соавторства применительно к исследованиям по физическим наукам в России за 2012–2018 годы. Показано, что ключевые глобальные свойства графа соавторства формируются публикациями с относительно небольшим количеством авторов. На основе анализа характеристик распределений степеней и взвешенных степеней вершин показано, что взвешенный граф с весами ребер, равными фракционному баллу, наиболее точно описывает основные особенности взаимодействия исследователей в данной области.

Ключевые слова: граф соавторства, физические науки, топология, распределение степеней вершин.

В текущей работе рассматриваются топологические свойства графа соавторства применительно к исследованиям по физическим наукам в России за 2012–2018 годы, построенного на основе информации о публикационной активности российских исследователей из базы данных WebOfScience. Глобальные характеристики данного графа изучены в работе [1], в настоящей работе более подробно изучаются ключевые локальные характеристики данного графа, а также оценивается роль коллаборационных статей в формировании глобальных свойств графа. Подробный обзор области, связанной с исследованием процесса генерации знаний посредством анализа сложных сетей, представлен в работе [2]. Для описания этого процесса используются сложные сети разного рода. Так, например, в работе [3] изучаются сети цитирования авторов, в работе [4] двудольные графы соавторства, а в работе [5] сеть цитирования между журналами. В настоящей работе рассматривается сеть, узлами которой являются авторы статей, а

¹ ФИАН, 119991 Россия, Москва, Ленинский пр-т, 53; e-mail: leonidovav@lebedev.ru.

² МФТИ, 141701 Россия, Московская обл., Долгопрудный, Институтский пер., 9.

ребра между данными узлами проводятся в том случае, если авторы написали хотя бы одну статью в соавторстве. Отличие работы [1] и данной работы от исследований, представленных в литературе, состоит, во-первых, в том, что впервые изучается российская сеть соавторства, построенная на данных о публикациях WebOfScience, а во-вторых, в том, что предложен принципиально новый способ определения весов ребер в данном графе.

Сетью соавторства \mathcal{G}_t называется граф, вершинами которого являются индивидуальные исследователи¹, а ребра между вершинами возникают при наличии хотя бы одной статьи, написанной данными исследователями в соавторстве за некоторый рассматриваемый период времени (t), например, год. Интенсивность и/или качество взаимодействия авторов могут быть учтены посредством добавления весов на ребра графа \mathcal{G}_t . Методика модификаций такого рода описана ниже.

Далее будут использованы следующие обозначения:

1. t – год, за который рассматривается граф (в данных $t = 2012, \dots, 2018$);
2. R_t – количество вершин графа \mathcal{G}_t , т. е. количество исследователей, являющихся соавторами хотя бы одной статьи за год t ;
3. $\mathcal{R}_t = \{\text{id}_1^t, \dots, \text{id}_{R_t}^t\}$ – множество вершин графа \mathcal{G}_t ;
4. \mathcal{G}_t – невзвешенный граф соавторства, ребро в котором демонстрирует наличие хотя бы одной совместной публикации у исследователей в году t ;
5. $\mathcal{G}_t = (\mathcal{R}_t, \mathcal{E}_t)$, где \mathcal{E}_t – множество ребер графа \mathcal{G}_t ;
6. матрица $\mathbf{G}_t = \{g_{ij}^t\}$, $i, j = 1, \dots, R_t$, – матрица смежности графа соавторства \mathcal{G}_t , в ней $g_{ij}^t = 1$, если исследователи id_i^t и id_j^t имели хотя бы одну общую публикацию в году t , и $g_{ij}^t = 0$, если это не так;
7. \mathcal{N}_t – взвешенный граф соавторства, в котором вес ребра равен количеству совместных статей соединяемых им узлов в году t ;
8. матрица $\mathbf{N}_t = \{n_{ij}^t\}$, $i, j = 1, \dots, R_t$, – матрица весов графа \mathcal{N}_t , для которой n_{ij}^t равно количеству совместных публикаций в году t у исследователей id_i^t и id_j^t ;
9. \mathcal{W}_t – взвешенный граф соавторства, в котором вес ребра равен сумме фракционных баллов совместных статей в году t ;
10. матрица $\mathbf{W}_t = \{w_{ij}^t\}$, $i, j = 1, \dots, R_t$, – матрица весов графа \mathcal{W}_t , для которой w_{ij}^t

¹В граф входят только те исследователи, у которых были работы в соавторстве с другими исследователями. Если исследователь публиковал статьи только в одиночку, то он не появится в графе и эти статьи никак не будут учтены.

расчитывается по следующей формуле:

$$w_{ij}^t = \sum_{k=1}^{n_{ij}^t} \frac{2Q_k}{\nu_k}, \quad (1)$$

где Q_k – это вес статьи k (зависящий от качества журнала, в котором опубликована данная статья, см. табл. 1), а ν_k – это количество авторов у статьи k .

Т а б л и ц а 1

Таблица весов статей в формуле (1)

Q1	Q2	Q3	Q4	Q	S	R	V	B
19.7	7.3	2.7	1	1	1	0.75	0.5	1

В таблице использованы следующие обозначения для изданий: Q1–Q4 – индексируемые Web of Science (WoS), Q – без кватриля из Web of Science Core Collection (WoSCC), S – индексируемые в Scopus, но не в WoS и WoSCC, R – индексируемые в RSCI WoS, но не в WoSCC, V – входящие в список ВАК, B – зарегистрированные в Российской книжной палате.

Ключевой локальной характеристикой графа является распределение степеней вершин. Степень d_i^t узла i в графе периода t определяется следующим соотношением

$$d_i^t = \sum_{j=1}^{R_t} g_{ij}^t. \quad (2)$$

В ходе анализа было установлено, что распределение степеней вершин практически не меняется во времени. А именно, в каждом периоде это распределение характеризуется наличием тяжелого хвоста, который хорошо описывается степенным распределением. Ввиду относительной статичности распределения, оценка параметра наклона степенного закона, приближающего хвост эмпирического распределения, производилась на объединенной за весь период анализа (с 2012 по 2018 г.г.) выборке. На рис. 1 изображено дополнение к кумулятивной функции распределения² (CCDF) степеней вершин.

²Дополнение кумулятивной функции распределения (Complimentary cumulative distribution function, CCDF) определяется следующим соотношением $CCDF(x) = 1 - F(x)$, где $F(x)$ – это кумулятивная функция распределения. Если хвосты распределения степенные, то $CCDF(x) \sim x^\alpha$, $x > x_m$, $\alpha < 0$, где x_m – параметр распределения. В связи с этим, в log-log масштабе $\ln CCDF(x) = \text{intercept} + \alpha \ln x$ для $x > x_m$. Таким образом, α может быть оценено как параметр наклона линейной регрессии.

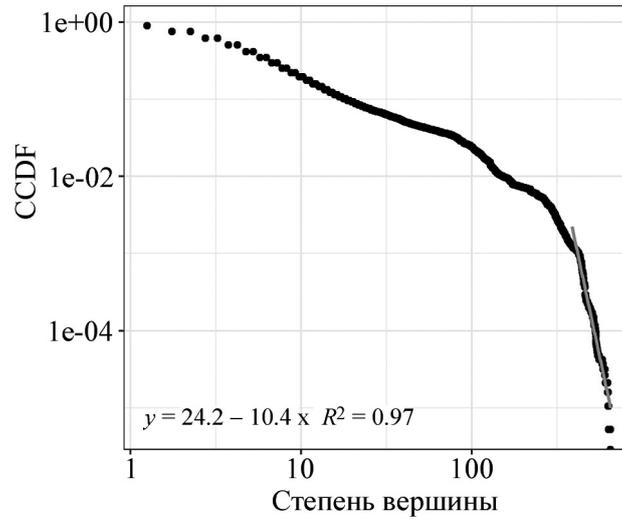


Рис. 1: Дополнение к кумулятивной функции распределения степеней вершин. Серая линия демонстрирует приближение степенным законом, параметры которого приведены на рисунке.

Для взвешенных графов \mathcal{N}_t и \mathcal{W}_t можно также определить взвешенные степени узлов, $s_i^{(n),t}$ и $s_i^{(f),t}$, $i = 1, \dots, R_t$, соответственно, следующим образом:

$$s_i^{(n),t} = \sum_{j=1}^{R_t} n_{ij}^t, \quad (3)$$

$$s_i^{(f),t} = \sum_{j=1}^{R_t} w_{ij}^t. \quad (4)$$

Аналогично распределению степеней вершин, распределения взвешенных степеней $s_i^{(n),t}$ и $s_i^{(f),t}$ относительно стабильны во времени, что позволяет оценивать параметры данных распределений по объединенной за период с 2012 по 2018 г.г. выборке наблюдений. CCDF данных распределений приведены на рис. 2.

Анализ диаграмм, представленных на рис. 1 и 2, показывает, что распределения степеней вершин и взвешенных степеней в графах \mathcal{N}_t характеризуются двумя близкими к линейным участками с существенно разным наклоном. При этом коэффициенты наклона степенного распределения для хвостов распределений оказываются достаточно большими по абсолютной величине (-10.4 и -18.3), что нехарактерно для большинства социальных сетей. Однако в графе \mathcal{W}_t хвост распределения описывается степенным законом с существенно меньшим по абсолютной величине угловым коэффициентом, равным -3.8 .

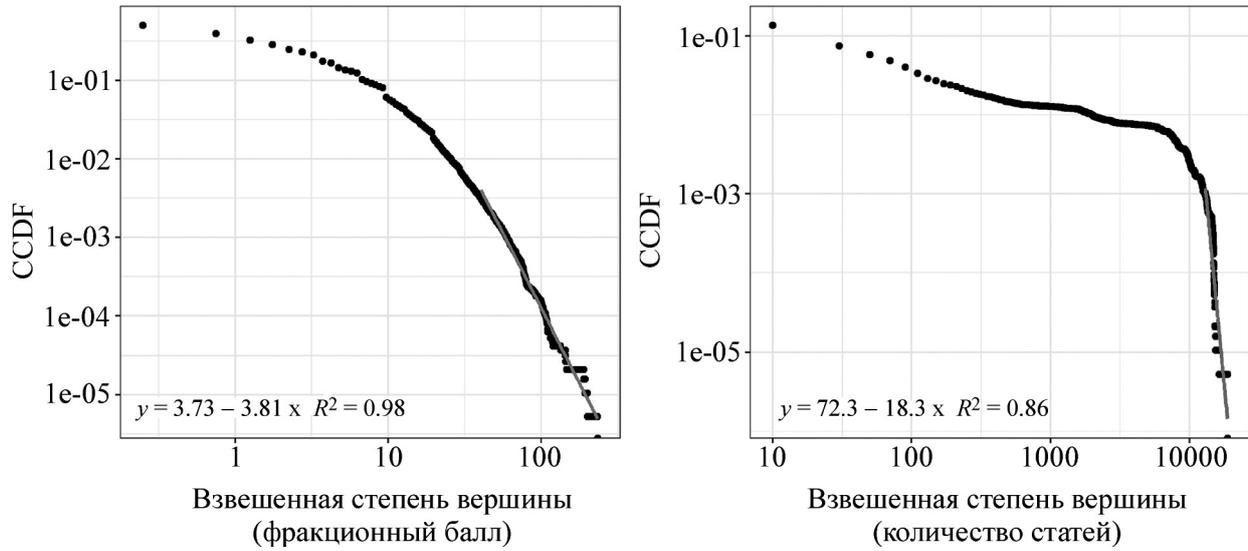


Рис. 2: Дополнения к кумулятивным функциям распределения взвешенных степеней вершин графов \mathcal{W}_t , $t = 2012, \dots, 2018$ (слева); \mathcal{N}_t , $t = 2012, \dots, 2018$ (справа). Серые линии демонстрируют приближение степенным законом, параметры которого приведены на рисунках.

Описанная особенность распределения степеней и взвешенных степеней в графе \mathcal{N}_t является следствием наличия в области физики больших коллабораций исследователей. Статьи с большим количеством авторов вносят существенный вклад в распределения степеней и взвешенных степеней в графе \mathcal{N}_t , однако имеют малый вклад в веса ребер графа \mathcal{W}_t .

Т а б л и ц а 2

Соотношение между глобальными характеристиками графов

\mathcal{G}_t и $\mathcal{G}_t(20)$, $t = 2012, \dots, 2018$

Год	R_t	$\frac{R_t(20)}{R_t}$	$ LCC_t $	$\frac{ LCC_t(20) }{ LCC_t }$
2012	18638	90%	9208	72%
2013	19093	92%	9294	70%
2014	21698	93%	11959	80%
2015	27440	94%	16733	87%
2016	31971	94%	20214	88%
2017	36097	94%	20975	88%
2018	34364	95%	20353	87%

Отметим, что коллаборационные публикации не вносят значительного изменения в глобальные характеристики графов. А именно, пусть $\mathcal{G}_t(20)$, $t = 2012, \dots, 2018$ – невзвешенный граф соавторства, построенный на основе информации о публикациях с не более чем 20 авторами, $R_t(20)$ – количество вершин в графе $\mathcal{G}_t(20)$. Обозначим также LCC_t множество вершин, входящих в наибольшую связную компоненту графа \mathcal{G}_t , и $LCC_t(20)$ – множество вершин, входящих в наибольшую связную компоненту графа $\mathcal{G}_t(20)$. В табл. 2 приведено соотношение между количеством вершин и размерами наибольших компонент графов \mathcal{G}_t и $\mathcal{G}_t(20)$. Из данной таблицы следует, что для каждого из рассматриваемых периодов почти все вершины графа \mathcal{G}_t (более 90%) присутствуют в графе $\mathcal{G}_t(20)$. И, кроме того, наибольшая компонента графа $\mathcal{G}_t(20)$ практически совпадает с наибольшей компонентой графа \mathcal{G}_t . Отсюда можно сделать вывод, что “каркас” графа соавторства формируется статьями с относительно небольшим количеством авторов. Иными словами, вероятно, большие коллаборации преимущественно формируются на основе уже существующих связей между научными сотрудниками, возникшими при работе в малых группах. Это же утверждение подтверждают результаты оценки параметров хвостов распределений размеров малых компонент, приведенные в первой строке табл. 3. Из данных результатов следует, что распределение размеров малых компонент практически не меняется при добавлении/исключении статей с большим количеством авторов.

Т а б л и ц а 3

Параметры, определяющие форму функций распределения характеристик в их хвостах. x_m – минимальное значение, относимое к хвосту распределения, α – угловой коэффициент CCDF соответствующих величин на графике в log-log масштабе. Под размером малых компонент понимается количество вершин в каждой из связных компонент графа, кроме наибольшей компоненты соответствующего графа (LCC_t или $LCC_t(20)$)

	Граф \mathcal{G}_t		Граф $\mathcal{G}_t(20)$	
	x_m	α	x_m	α
Размер малых компонент	17	-2.8	13	-2.4
d_i^t	500	-10.4	40	-5.5
$s_i^{(n),t}$	10000	-18.3	100	-6.0
$s_i^{(f),t}$	40	-3.8	50	-5.0

Сравнение оценок параметров распределений степеней (d_i^t) и взвешенных степеней ($s_i^{(n),t}, s_i^{(f),t}$) графа \mathcal{G}_t (\mathcal{N}_t и \mathcal{W}_t) и графа $\mathcal{G}_t(20)$ ($\mathcal{N}_t(20)$ и $\mathcal{W}_t(20)$) показывает, что распределение d_i^t и $s_i^{(n),t}$, действительно, крайне чувствительно к добавлению коллаборационных статей, тогда как на распределение $s_i^{(f),t}$ такие статьи существенного влияния не оказывают. Исходя из этого можно сделать вывод, что свойства графа \mathcal{W}_t с весами, равными фракционным баллам, более корректно характеризуют взаимодействие исследователей в области физики в России, чем его невзвешенный аналог \mathcal{G}_t и взвешенная модификация \mathcal{N}_t .

В заключение перечислим основные результаты, описанные в настоящей работе. При рассмотрении графа соавторства работ в области физики в России за 2012–2018 г.г. показано, что:

- глобальные свойства графа соавторства, такие как количество вершин и компонентная структура, формируются публикациями с относительно небольшим количеством авторов;
- хвосты распределений степеней вершин графа соавторства и взвешенных степеней двух его взвешенных модификаций могут быть приближены степенным законом;
- распределение степеней в графе соавторства и взвешенных степеней вершин в его модификации с весами, равными количеству совместных статей у авторов, значительно меняется при исключении коллаборационных статей;
- распределение взвешенных степеней вершин в графе с весами, равными фракционному баллу, устойчиво к добавлению/исключению коллаборационных статей и, тем самым, данный граф наиболее точно описывает основные особенности взаимодействия исследователей.

Статья опубликована в рамках проекта “Посткризисное мироустройство: вызовы и технологии, конкуренция и сотрудничество” по гранту Министерства науки и высшего образования РФ на проведение крупных научных проектов по приоритетным направлениям научно-технологического развития (соглашение № 075-15-2020-783).

ЛИТЕРАТУРА

- [1] О. В. Иванов, А. М. Коваленко, А. В. Колобов и др., Краткие сообщения по физике ФИАН **47**(8), 9 (2020). DOI: 10.3103/S1068335620080060.

- [2] A. Zeng, Z. Shen, J. Zhou, et al., *Physics Reports*, No. 714, 1 (2017). DOI: 10.1016/j.physrep.2017.10.001.
- [3] S. Lehmann, B. Lautrup, A. D. Jackson, *Phys. Rev. E* **68**(2), 026113 (2003). DOI: 10.1103/PhysRevE.68.026113.
- [4] Michel L. Goldstein, Steven A. Morris, Gary G. Yen, *Phys. Rev. E* **71**(2), 026108 (2005). DOI: 10.1103/PhysRevE.71.026108.
- [5] Vicente P. Guerrero-Bote, Félix Moya-Anegón, *Journal of Informetrics* **6**(4), 674 (2012). DOI: 10.1016/j.joi.2012.07.001.

Поступила в редакцию 27 октября 2020 г.

После доработки 30 октября 2020 г.

Принята к публикации 30 октября 2020 г.