

О СООБЩЕСТВАХ СОАВТОРСТВА УЧЁНЫХ (НА ПРИМЕРЕ РОССИЙСКИХ МАТЕМАТИЧЕСКИХ ЖУРНАЛОВ)

А.А. Печников

*Институт прикладных математических исследований – обособленное подразделение ФГБУН ФИЦ
«Карельский научный центр Российской академии наук»
Россия, 185910, Петрозаводск, ул. Пушкинская, 11
E-mail: pechnikov@krc.karelia.ru*

Ключевые слова: наукометрия, соавторство, граф, модулярность, сообщество.

Аннотация: Предложена методология исследования тематических сообществ, основанная на графе соавторства ученых, работающих в рамках одной научной дисциплины. Разбиение графа на тематические сообщества позволяет получить базовые характеристики сообществ, такие как их тип (тематическое направление исследований), количество ученых, входящих в каждое сообщество и связей между ними. Анализ построенного разбиения даёт перечень основных направлений исследований в рамках научной дисциплины и обобщенные данные типа суммарного количества сообществ и ученых в каждом научном направлении, что, в свою очередь, может быть использовано для принятия управленческих решений по стимулированию по направлениям, представляющими безусловную актуальность и практическую ценность на сегодняшний день. Изложены результаты апробации предложенной методологии исследования на основе данных математического портала Math-Net.Ru.

1. Введение

Анализ наукометрической информации тесно связан с задачами прогноза развития науки, повышения качества существующих, выявления и развития новых направлений исследований. Изучение научных связей дает возможность оценивать тенденции развития различных научных направлений, выявлять научные центры, научные школы и значимые персоны.

Методы и подходы к решению таких задач во многом зависят от масштаба исследования. К задачам мирового масштаба можно отнести построение глобальной карты мировой науки Map of Science [1]. На сегодня она включает в себя более 130 миллионов научных публикаций со всего мира, алгоритмически организованных в более чем 116 тысяч исследовательских кластеров.

Основным ограничением для оценки научной деятельности российских организаций с использованием евроатлантических возможностей является то, что значительная часть научных результатов публикуется в российских журналах, многие из которых не попадают в базы Scopus и Web of Science и не цитируются зарубежными учеными.

Кроме того, масштабные проекты создаются по единому методу для различных научных направлений, что не всегда удобно для анализа отдельных научных дисциплин, теряющихся на визуальных представлениях, возможной неполноты данных и высоких барьеров цитирования и/или соавторства в масштабных проектах.

Известно большое количество российских и зарубежных работ, основанных на библиографической информации, в которых исследуются коммуникации ученых в

различных областях науки и затрагиваются различные аспекты этой проблематики. Например, в [2] подход, основанный на исследовании динамики графа соавторства с использованием собственной базы данных, используется для изучения российских региональных групп по истории. В [3] по данным за последние шестьдесят лет показано, что вероятность написания в соавторстве статьи в журналах по общей экономике выше, чем в журналах по экономической истории. Более традиционные работы в области физики [4] и экономики [5] посвящены исследованиям характеристик и свойств графов соавторства.

Данная работа посвящена попыткам разработки методологии исследования взаимодействия ученых, понимаемого как коммуникации между представителями различных разделов одной научной дисциплины, реализуемые посредством соавторства, то есть создания общими усилиями публикаций, основанных на совместном решении определенных задач.

2. Методология исследования тематических сообществ соавторства ученых

Математической моделью, представляющей объект нашего исследования, является граф соавторства, вершины которого соответствуют ученым-авторам публикаций, а ребра между вершинами отражают связи между соавторами. Как в [6]: «...два ученых считаются связанными, если они совместно написали статью», поскольку «... большинство людей, которые написали статью вместе, будут хорошо знать друг друга». То есть связь между учеными не зависит от количества написанных совместно статей, – она есть, если есть хотя бы одна такая статья. Графы соавторства являются одними из часто исследуемых конструкций научного сотрудничества из-за очевидности определения «знакомства» учёных, что документируется самими статьями. Будем исходить из того, что исследователи располагают данными о некотором множестве ученых и их публикационной активности за определенный промежуток времени, позволяющими построить граф соавторства.

Формально граф соавторства это $G=G(V,E)$, где V – множество вершин, соответствующих авторам, а E – множество ребер, соединяющих пары авторов, написавших в соавторстве хотя бы одну статью. В связи со сказанным ранее $G(V,E)$ – неориентированный граф без петель и кратных ребер.

Характерная особенность реальных сетей заключается в наличии у них свойства кластеризации, в соответствии с которым топология графа организована в сообщества (также называемые модулями, группами или кластерами), т.е. подграфы, имеющие больше связей внутри себя, чем вовне [6].

Пусть k – это количество сообществ, на которые разбивается граф соавторства $G=G(V,E)$, а V_0, V_1, \dots, V_{k-1} – подмножества вершин, составляющих сообщества. Любая пара подмножеств не пересекается, т.е. $V_i \cap V_j (i,j=0, \dots, k-1; i \neq j)$, а их объединение даёт множество $V: V_0 \cup V_1 \cup \dots \cup V_{k-1} = V$.

Одно из часто используемых на практике формальных определений меры модулярности Q , которое мы используем, дано в [7]. Значение Q при некотором заданном разбиении на сообщества вычисляется как доля рёбер от общего числа рёбер, которые попадают в данные сообщества минус ожидаемая доля рёбер, которые попали бы в те же сообщества, если бы они были распределены случайно. При таком определении значение $Q \in [-1, 1]$ и разбиение считается «хорошим», если значение $Q > 0,7$.

Одним из основных допущений, используемых далее, является допущение о том, что авторы, попадающие в одно сообщество, в основном работают над некоторой

единой тематикой, определяющей научное направление данного сообщества. Таким образом, встает задача определения тематики (научного направления, научной дисциплины) группы ученых по их совместным публикациям. Такая формулировка сама порождает задачу, требующую отдельной постановки и исследования. Методы ее решения во многом зависят от масштаба исследования. Очевидно, что в каждом конкретном случае будут использоваться свои классификаторы научных направлений и методы для определения принадлежности ученых к тому или иному направлению.

Например, далее для классификации математических сообществ будет взята номенклатура научных специальностей, а тематика сообществ будет определяться по номенклатуре наиболее авторитетных ученых, входящих в эти сообщества.

В результате получаем разбиение графа $G(V, E)$ на сообщества C_0, C_1, \dots, C_{k-1} , где с каждым сообществом C_i связана тройка (V_i, E_i, t_i) , а именно, множество вершин, множество связывающих их ребер и тематика сообщества.

Это позволяет провести анализ результатов с целью вычисления базовых характеристик тематических сообществ, таких как количество ученых, входящих в каждое сообщество и связей между ними, количество связей между сообществами и т.д.

Не исключено также, что некоторые сообщества будут иметь одинаковую тематику, что позволит получить обобщенные данные типа суммарного количества ученых в каждом укрупненном научном направлении.

Содержательная интерпретация полученных результатов и выводы существенно зависят от масштаба задачи и исследуемых научных дисциплин, и далее выполнена нами применительно к проведенному анализу соавторства российских математиков.

3. Исследование тематических сообществ соавторства математиков

В информационной системе Math-Net.Ru данные накапливаются в течение последних 15 лет. Наполнение информационной системы по годам было неоднородным, поэтому для исследования был выбран временной интервал датировки статей, опубликованных с 2000 по 2020 год; из 140 журналов были взяты 56 ведущих российских журналов, полностью индексируемые в Math-Net.Ru. В список 56 журналов входят «Алгебра и анализ», «Математические заметки», «Сибирский математический журнал» и др. Подробная информация о журналах приводится в [8, с. 194]. С 2000 по 2020 год в этих журналах опубликовано почти 62 тысяч статей, авторами которых являются около 33 тысяч ученых.

Была сделана выборка всех авторов, имеющих публикации в соавторстве за указанный период в любом из 56 журналов и на основе этого сформировано множество вершин графа V . Множество ребер E формируется на основе выборки из всех возможных пар соавторов по правилу: для любой пары $i, j \in V$ ребро $(i, j) \in E$ существует тогда и только тогда, когда авторы i и j имеют хотя бы одну статью, в которой они входят в состав соавторов.

Далее из графа $G(V, E)$ был взят подграф, равный максимальной компоненте связности (11860 вершин и 20858 ребер), и несколькими последовательными итерациями из него были удалены вершины с инцидентностью 1. Удаление таких вершин в дальнейшем не оказало влияния на связность графа, но существенно облегчило работу по содержательному анализу сообществ.

В итоге был построен граф соавторства, обозначаемый $G_{co-auth}$, состоящий из одной компоненты связности, содержащий 8166 вершин и 17164 ребер, и являющийся объектом дальнейших исследований. Диаметр графа равен 26, а средняя длина пути 9,7.

Для разбиения графа $G_{co-auth}$ на сообщества был использован пакет Gephi [9]. Максимальное значение меры модулярности Q , равное 0,933, достигается при разбиении графа на 74 сообщества.

Для определения типа каждого сообществ использовался следующий подход. Из множества авторов, составляющих сообщество, отбирались первые три автора, имеющие наибольшее количество соавторов в данном сообществе, а затем по информации Math-Net.Ru в качестве типа сообщества принималась номенклатура специальности ВАК этих авторов. Номенклатура специальностей не является обязательным атрибутом в базе данных авторов, и тогда приходилось исследовать дополнительную информацию, например, номенклатуру журналов, в которых есть публикации данных авторов. Использование старой номенклатуры [10] объясняется тем, что у многих авторов в Math-Net.Ru указана именно она.

Например, в сообществе, обозначаемом C_{38} и содержащим 94 автора и 137 ребер, первая тройка авторов – это академик Д.А. Новиков (19 соавторов, специальность ВАК 05.13.10), д.т.н. В.Н. Бурков (16, специальность докторской диссертации 1975 года 05.00.00) и д.т.н. В.А. Иващенко (14). Поэтому тип сообщества C_{38} принят как, 05.13.10 – управление в социальных и экономических системах.

4. Некоторые результаты

В таблице 1 приведены сводные результаты по сообществам, имеющим одинаковую научную специальность.

Шесть научных направлений, по которым работа ведется как наибольшим количеством сообществ, так и максимальным суммарным количеством их участников, в таблице выделены более толстыми линиями ячеек с названиями специальностей.

Три научных направления находящихся в очевидных аутсайдерах по количеству участников и сообществ, выделены в таблице свето-серой заливкой.

Таблица 1. Сводная таблица сообществ графа $G_{co-auth}$.

Шифр	Наименование научной специальности	Кол-во сообществ	Кол-во ученых	Среднее кол-во
01.01.01	вещественный, комплексный и функциональный анализ	7	692	98,9
01.01.02	дифференциальные уравнения, динамические системы и оптимальное управление	8	746	93,3
01.01.03	математическая физика	9	907	100,8
01.01.04	геометрия и топология	4	478	119,5
01.01.05	теория вероятностей и математическая статистика	2	358	179,0
01.01.06	математическая логика, алгебра и теория чисел	7	955	136,4
01.01.07	вычислительная математика	10	1036	103,6
01.01.09	дискретная математика и математическая кибернетика	3	344	114,7
01.02.01	теоретическая механика	2	323	161,5
01.02.04	механика деформируемого твердого тела	3	340	113,3
01.02.05	механика жидкости, газа и плазмы	3	230	76,7
01.03.01	астрометрия и небесная механика	1	135	135,0
01.04.02	теоретическая физика	3	324	108,0
01.04.17	химическая физика, горение и взрыв, физика экстремальных состояний вещества	1	36	36,0
05.02.05	роботы, мехатроника и робототехнические системы	1	47	47,0

05.13.10	управление в социальных и экономических системах	2	220	110,0
05.13.16	применение вычислительной техники, математического моделирования и математических методов в научных исследованиях	2	232	116,0
05.13.18	математическое моделирование, численные методы и комплексы программ	5	742	148,4
05.13.19	методы и системы защиты информации, информационная безопасность	1	21	21,0

С практической точки зрения, полученные результаты могут быть непосредственно использованы для обоснования управленческих решений с целью стимулирования работ по таким выявленным направлениям, как роботы и робототехнические системы, горение и взрыв, методы и системы защиты информации, представляющими безусловную актуальность и практическую ценность на сегодняшний день.

Результаты апробации предлагаемого подхода показывает как потенциал его использования, так и возможностей переноса на другие научные дисциплины. Основным условием является наличие полной и достоверной библиографической информации о соавторстве ученых.

Автор приносит благодарность разработчикам проекта Math-Net.Ru за многолетнее успешное сотрудничество и предоставленные данные.

Список литературы

1. Map of Science. <https://sciencemap.eto.tech> (дата обращения: 04.01.2024).
2. Гарскова И.М. Сетевой анализ историографии: динамика формирования региональных центров исторической информатики // Историческая информатика. 2017. № 3. С. 94-115.
3. Seltzer A.J., Daniel S.H. Co-authorship in economic history and economics: Are we any different? // Explorations in Economic History. 2018. Vol. 69. P. 102-109.
4. Васильева Е.Е., и др. Влияние коллабораций на характеристики графа соавторства в области физики в России // Краткие сообщения по физике ФИАН. 2020. № 12. С. 47-54.
5. Бредихин С.В., Ляпунов В.М., Щербакова Н.Г. Ранжирование узлов взвешенной сети соавторства: анализ данных БД REPEC // Проблемы информатики. 2021. №. 4 (53). С. 67-83.
6. Newman M.E.J. The structure of scientific collaboration networks // Proceedings of the National Academy of Sciences of the USA. 2001. Vol. 98, No. 2. P. 404-409.
7. Newman M.E., Girvan M. Finding and evaluating community structure in networks // Physical Review E. 2004. Vol. 69, No. 2. P. 026113.
8. Знаменская Е.А., Печников А.А., Чебуков Д.Е. Анализ соавторства в математических журналах Math-Net.Ru // Научный сервис в сети Интернет: труды XXIV Всероссийской научной конференции (19-22 сентября 2022 г., онлайн). М.: ИПМ им. М.В. Келдыша, 2022. С. 190-202.
9. Gephi – The Open Graph Viz Platform. <https://gephi.org> (дата обращения: 03.01.2024).
10. Приказ Министерства науки и высшего образования РФ от 23 октября 2017 г. № 1027 «Об утверждении номенклатуры научных специальностей, по которым присуждаются ученые степени». <https://docs.cntd.ru/document/542610966> (дата обращения: 10.01.2024).