

МЕТОДИКА СЕТЕВОГО АНАЛИЗА НАУЧНЫХ ПУБЛИКАЦИЙ

И. Г. Ольгина^[0000-0002-9932-4552]

Омский государственный технический университет

inna_olgina@mail.ru

Аннотация

Актуальность вопросов анализа значимости научных публикаций обусловлена тем, что с появлением интернет-технологий стал возможен сбор данных о сети цитирования публикаций. Между тем, существующий сегодня подход к анализу значимости научных публикаций базируется на библиометрических показателях, учитывающих только количество цитирований. Однако все более широкое применение начинает получать сетевой анализ, применяемый преимущественно в исследованиях социальных сетей. Автором разработана методика, позволяющая осуществить эффективный анализ значимости научных публикаций, которая основана на методах сетевого анализа, альтернативных библиометрическим методам. В качестве критериев оценки значимости научных публикаций, основанных на сетевом анализе, установлены релевантные меры центральности узлов сети цитирования: центральность по степени связности; близости к другим узлам; посредничеству; авторитетности; концентрации. Приведен результат эксперимента, позволивший продемонстрировать адекватность разработанной методики анализа научных публикаций на основе сетевых метрик. В качестве первичных источников данных о публикациях использованы наукометрические базы данных, позволяющие отслеживать цитируемость публикаций и выявлять соответствующие сети цитирования. Применение предложенной методики способствует выявлению важных публикаций в развитии соответствующих научных направлений.

Ключевые слова: сеть цитирования, публикации, наукометрия, библиометрический анализ, сетевой анализ, граф.

ВВЕДЕНИЕ

С увеличением количества информации, представляемой в виде научных публикаций, появляется необходимость анализа разнообразных показателей уровня значимости этих источников, включая их информативность, авторитетность в научной среде, количества ссылок на них и т. д. На сегодняшний день существует очень большое количество международных систем цитирования (библиографических баз): РИНЦ, Web of Science, Scopus, Web of Knowledge, Chemical Abstracts, PubMed, Springer, Agris, GeoRef и др., которые используются для оценки уровня научных публикаций. При этом применяются признанные в мире библиометрические показатели, которые характеризуют авторов или журналы: импакт-фактор, индекс Хирша, индекс цитирования [1, 2, 3]. Но в расчете таких показателей и индексов учитывается только количество цитирований – для качественной оценки публикаций этого недостаточно.

Появление науки о сетях – Network Science позволяет исследовать сложные сетевые системы (в том числе социальные и информационные сети) посредством представления их в виде графовых моделей. Широкое применение приобрел метод сетевого анализа [4]. За последние десятилетия возрос интерес к науке о сетях, что повлекло за собой закономерное развитие всевозможных инструментов для исследований в данной области.

Рост в геометрической прогрессии общего объема публикаций обуславливает актуальность задач анализа взаимосвязей научных публикаций. В науке о сетях для решения этих задач разрабатываются модели и методы, относящиеся к сфере так называемых сетей цитирования. Однако оценить важность научных публикаций с учетом многоаспектного сетевого анализа не представляется возможным в силу отсутствия соответствующих инструментов. Исходя из этого, ставится задача комбинирования сетевых мер для выявления наиболее важных публикаций.

Существует различное программное обеспечение для визуализации и исследования сетей, например, VOSviewer (<https://www.vosviewer.com/>), Gephi (<https://gephi.org/>), Tom Sawyer Perspectives (<http://www.tomsawyer.com/>), Sentinel Visualizer (<http://www.fmsasg.com/Products/SentinelVisualizer/>). Эти инструменты широко применяются для исследования социальных сетей, в частно-

сти, сетей цитирования. Возможно построение сетей цитирования патентов [4], ключевых слов [5], научных публикаций [6], сетей авторского цитирования [7], сетей по совместному цитированию и соавторству [8, 9, 10]. Для анализа таких сетей разработано большое количество показателей (мер центральности), характеризующих значимость каждого узла в разных аспектах [11]. Исследование мер центральности представляет действительно большой интерес в исследованиях социальных сетей и сетей цитирования. По этой теме имеется достаточно много публикаций. В статье [12] дан обзор исследовательских работ по показателям центральности в социальных сетях.

На практике основной слабой стороной программных продуктов, названных выше, является отсутствие возможности определения суммарной оценки с учетом нескольких показателей. Это существенно ограничивает исследования сетей и визуализацию результатов с учетом многоаспектности. Таким образом, возникает потребность в решении задачи вывода и визуализации результатов многокритериальной оценки узлов сети.

Пример комбинирования методов для достижения лучшей производительности и получения более точных показателей по сравнению с традиционными методами продемонстрирован в работе [4]. Ее авторы объединили классический анализ основного пути (main path analysis) с алгоритмом PageRank и протестировали этот новый комбинированный метод на доступных данных о патентах. Стоит отметить, что анализ сети цитирования патентов является одним из наиболее важных методов измерения значимости интеллектуального анализа и идентификации содержания патентов.

ОПРЕДЕЛЕНИЕ ЗНАЧИМОСТИ ПУБЛИКАЦИИ НА ОСНОВЕ СЕТЕВЫХ МЕТРИК

Предметом настоящего исследования являются сети цитирования, которые представляют собой один из видов социальных сетей. Математическими моделями сетей цитирования являются ориентированные графы. Узлами сети являются научные публикации, а связями – коммуникации между ними, реализуемые путем цитирования [6]. Сеть цитирования научных публикаций представляется в виде ориентированного графа $G = (V, E)$, где V – множество вершин графа, а E – множество его дуг.

При вычислении значений параметров узлов сети, которыми являются меры центральности, удобно использовать ранжирование узлов сети по каждому из параметров в отдельности. Например, в статье [13] выполнено ранжирование коллекции периодических изданий на основе центральности по посредничеству и предложен алгоритм вычисления этой меры для взвешенных графов. В статье [14] подтверждена адекватность графа цитирования журналов цифровой библиотеки Math-Net.Ru как модели научных коммуникаций сравнением ранжирования журналов в графе цитирования с их рейтингом SCIENCE INDEX в eLIBRARY.RU. Ранжирование научных журналов осуществлялось по значению меры Page Rank.

По значениям мер центральности упорядочиваются по важности узлы сети при следующих ограничениях. Первое заключается в том, что центральность, которая оптимальна для одного приложения, часто не оптимальна для другого. Следовательно, не нужно использовать столько различных центральностей. Второе ограничение состоит в том, что центральность вершины отражает относительную важность вершин в графе. Меры центральности для измерения узлов в общем виде не разрабатывались [15].

При исследовании значимости публикации в соответствующей научной области с помощью сетевого подхода необходимо применить многокритериальный анализ. Следовательно, недостаточно оценить публикацию с помощью единственного показателя – одной из перечисленных мер центральности, а необходимо учесть важность соответствующего узла сети с помощью релевантных мер центральности C , исходя из целей исследования. Поэтому необходимо вычислить «обобщенный показатель важности», который представляет собой взвешенную сумму частных показателей C_i , в которую каждый из них входит с определенным весом k_i , отражающим его значимость:

$$C = k_1 C_1 + k_2 C_2 + \dots + k_n C_n, \quad (1)$$

где $0 \leq k_i \leq 1, i = \overline{1, n}; n \in \mathbf{N}$. Весовые коэффициенты, с которыми входят в расчет разные показатели, не постоянны, а изменяются в зависимости от ситуации.

МЕТОД ОПРЕДЕЛЕНИЯ ВАЖНОСТИ УЗЛА СЕТИ ЦИТИРОВАНИЯ

Данный метод можно также назвать методом комплексной оценки центральности узлов сети. В соответствии с представленным выше способом (формула (1)) строится математическая модель, которая позволяет определить комплексную оценку центральности узлов сети цитирования как сумму рангов узлов сети по каждой мере центральности с учетом их релевантности, по следующей формуле:

$$C_{sum}(h) = \sum_{i=1}^n k_i(h)R_i, \quad (2)$$

где R_i – ранг i -го показателя; n – число показателей; k_i – весовой коэффициент – индекс релевантности i -го показателя, $0 \leq k_i \leq 1$; h – профиль исследования.

Для удобства проведения экспериментов в настоящей работе индекс релевантности k , соответствующий весу показателя, принимается равным 1, если данный показатель соответствует профилю исследования, и 0, если он не принимается во внимание, т. е. $k_i \in \{0, 1\}$.

ОТБОР ПОКАЗАТЕЛЕЙ ДЛЯ АНАЛИЗА СЕТЕЙ ЦИТИРОВАНИЯ НАУЧНЫХ ПУБЛИКАЦИЙ

Нами проведен анализ существующих мер центральностей узлов для социальных сетей. На этой основе выбраны пять показателей для анализа сетей цитирования: центральность по степени связности (degree centrality) [16]; центральность по близости к другим узлам (closeness centrality) [17]; центральность по посредничеству (betweenness centrality) [18]; центральность по авторитетности (authority centrality) и центральность по информативности (hub centrality) [19, 20]. В работе [21] приведено подробное описание этих мер центральности узлов сети. В [15] дана интерпретация значений каждой меры центральности относительно того, как она может повлиять на показатель важности публикации. Для проведения исследований по обнаружению зарождения инноваций в определенной научной области с помощью сетевого анализа применяют центральность по близости к другим узлам и центральность по посредничеству. Примером могут служить исследования [22]. Для исследования социальных сетей в целом применяют центральность по посредничеству [23]. Центральность по авто-

ритетности в социальных сетях применяют для рекомендаций пользователю, на кого стоит ориентироваться, как продемонстрировано в [24]. Если использовать этот показатель в контексте сетей цитирования, то можно интерпретировать его так: данная статья располагает важной информацией по определенной теме. Центральности по авторитетности и информативности фокусируются на структуре сети и определяют ее важность в соответствии с их позициями на графе [25] и зависят от общего количества отношений, которые имеются за пределами узла [24]. Узел является центром, когда он имеет ребра с авторитетными узлами.

Является важным, что в связи с тем, что граф сети цитирования является ориентированным, входящие и исходящие связи можно анализировать отдельно. При определении центральности по близости к другим узлам для орграфов будем рассматривать как дистанции от определяемой вершины до всех остальных (исходящие связи – режим *out*), так и дистанции от всех вершин до определяемой (входящие связи – режим *in*) [15]. В случае, когда цитируемым объектам придается большая значимость, имеет смысл пользоваться вторым определением. При рассмотрении центральности по степени связности аналогично рассматриваются отдельно входящие связи – (полустепень захода), которые соответствуют количеству цитирований рассматриваемой публикации, и исходящие связи (полустепень исхода), отражающие количество ссылок публикации на другие.

Стоит отметить, что при анализе социальных сетей часто применяют показатель PageRank. Эта центральность изначально применялась поисковой системой Google для ранжирования веб-страниц и рассчитывается по формуле

$$PR(A) = (1 - d) + d \sum_{i=1}^n \frac{PR(T_i)}{C(T_i)}, \quad (3)$$

где A – оцениваемая веб-страница, T_i – веб-страницы, ссылающиеся на страницу A , d – коэффициент демпфирования (вероятность перехода по ссылке, имеющейся на странице A), $C(T)$ – число ссылок на веб-странице T .

В случае анализа научных публикаций эту метрику можно не принимать во внимание в силу особенности данной меры и целей, для которых она создавалась. В формуле (3) расчета значений названной меры применяется коэффициент демпфирования d , который в сети цитирования может быть проинтерпретирован как

вероятность того, что читатель просмотрит источник, ссылка на который есть в публикации.

МЕТОДИКА СЕТЕВОГО АНАЛИЗА НАУЧНЫХ ПУБЛИКАЦИЙ

Предложим новую методику ранжирования публикаций на основе сетевых метрик, комбинация которых задает профиль ранжирования. Этот профиль определяется целью исследования.

1. Сбор данных о цитировании

Для проведения анализа научных публикаций необходимы данные о цитировании, которые можно получить следующими способами:

- сбор данных с использованием покластерного обхода всей сети цитирования (с помощью написанной программы);
- сбор подсети по заданной тематике;
- использование данных о сети цитирования публикаций определенной научной организации; например, в Омском государственном техническом университете создан с сервис science.omgtu.ru, содержащий необходимые данные для сетевого анализа публикаций всех авторов университета.

2. Выбор и формирование профиля исследования

Нами рассмотрены меры центральности узлов сети, которые следует использовать в качестве характеристик узлов сети цитирования научных публикаций.

Для математического описания задачи введем следующие обозначения:

$K = \{k_1, k_2, \dots, k_n\} (n = 9)$ – кортеж коэффициентов, задающих степень влияния соответствующей центральности на профиль исследования (индекс релевантности i -го показателя);

$H = \{h_1, h_2, \dots, h_m\}$ – множество профилей исследования, которые формируются, исходя из поставленных целей. К примеру, если важность публикации характеризуется ее высокой информативностью, то такие публикации будут содержать большое количество ссылок или иметь ссылки в основном на авторитетные публикации.

Для характеристики узлов сетей цитирования используем следующие сетевые метрики:

- k_1 – центральность по степени связности (degree centrality);

- k_2 – полустепень исхода (out-degree centrality);
- k_3 – полустепень захода (in-degree centrality);
- k_4 – центральность по близости к другим узлам (closeness centrality);
- k_5 – центральность по близости к другим узлам (closeness centrality в режиме *out*);
- k_6 – центральность по близости к другим узлам (closeness centrality в режиме *in*);
- k_7 – центральность по посредничеству (betweenness centrality);
- k_8 – центральность по авторитетности (authority centrality);
- k_9 – центральность по информативности (hub centrality).

Таким образом, k_i является весовым коэффициентом и принадлежит множеству рациональных чисел, $0 \leq k_i \leq 1, i = \overline{1, n}$.

Выбор значения k_i зависит от предпочтений лица, принимающего решение (ЛПР). Для простоты проведения экспериментальных исследований рекомендуется использовать $k_i = 1$, если показатель принимается во внимание, и $k_i = 0$ в противоположном случае.

Профиль исследования h_j формируется с учетом множества значений весовых коэффициентов K_j в виде кортежа (k_1, k_2, \dots, k_i) , $i = 9$, где $j = 9^2$ при $k_i \in \{0, 1\}$. Соответственно количество профилей исследования будет составлять 81 вариант.

ЛПР необходимо определить сочетание показателей, которые будут составлять профиль исследования, а также определить весовые коэффициенты данных показателей.

Одним из подходов к решению таких задач является принятие решения на основе методов ранговой корреляции Спирмена и Стьюдента [26, 27]. Методы ранговой корреляции могут быть использованы для определения тесноты связи не только между количественными, но и качественными переменными при условии, если их значения можно ранжировать и упорядочить. Согласно этим методам значения комплексной оценки узлов сети, полученные по формуле (2), сравниваются с известными альтернативными показателями, применяемыми в базах цитирования. Затем требуется выбрать профиль исследования, при кото-

ром значения коэффициента корреляции наибольшие, при условии отклонения гипотезы H_0 об отсутствии зависимости.

3. Алгоритм ранжирования научных публикаций по важности

Алгоритм, реализующий метод ранжирования публикаций по важности согласно профилю исследования на основе сетевого анализа, представлен на рисунке 1.

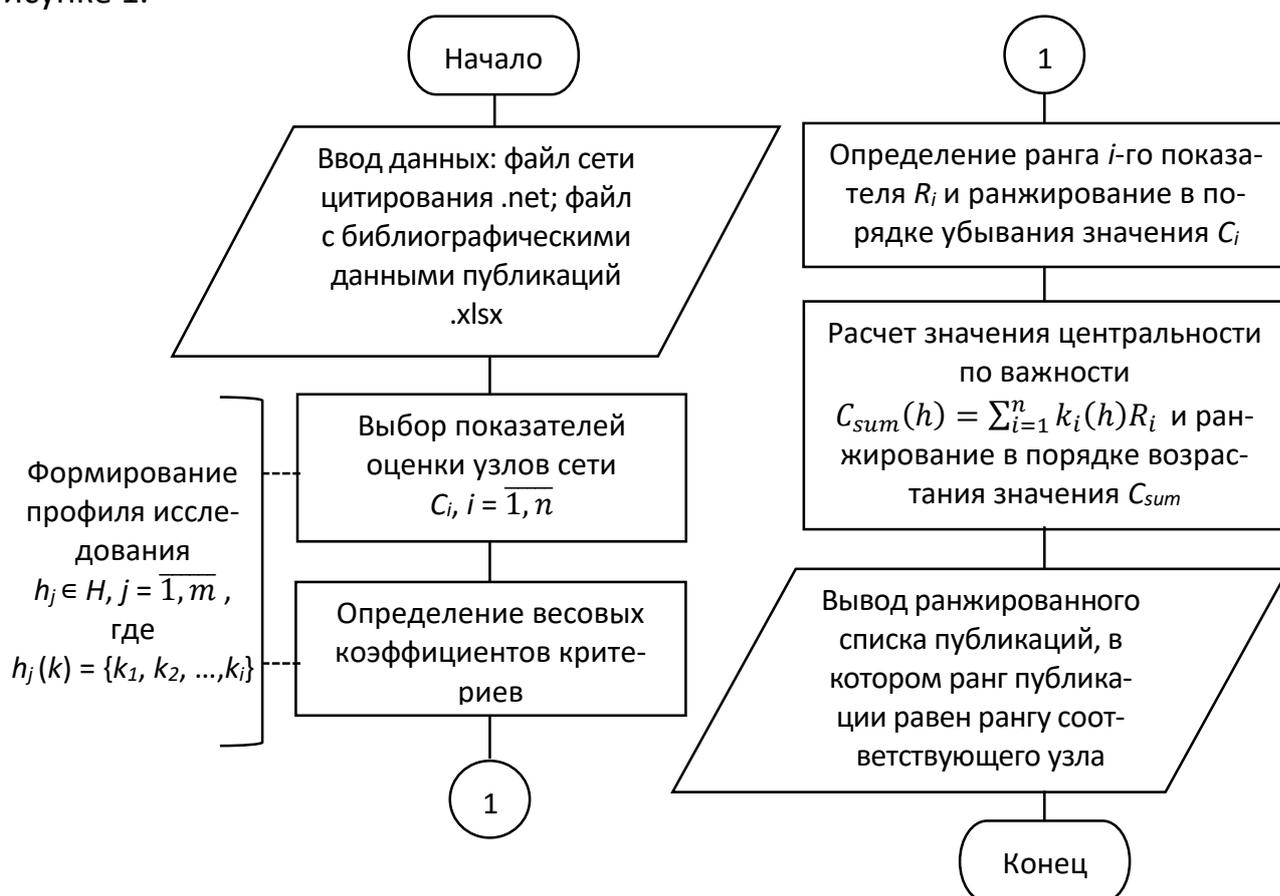


Рис. 1. Алгоритм ранжирования научных публикаций по важности

С целью реализации предложенной методики анализа сетей цитирования разработаны «Программный комплекс LinkAnalyzer 1.0 для сбора и анализа информации о цитировании научных публикаций» № 2020615709 от 29.05.2020 г.; «Генератор списка источников информации в сетях цитирования» № 2021661693 от 14.07.2021 г.; «Визуализатор сетей цитирования» № 2023666387 от 31.07.2023 г. Интерфейс программы для анализа научных публикаций представлен на рисунке 2.

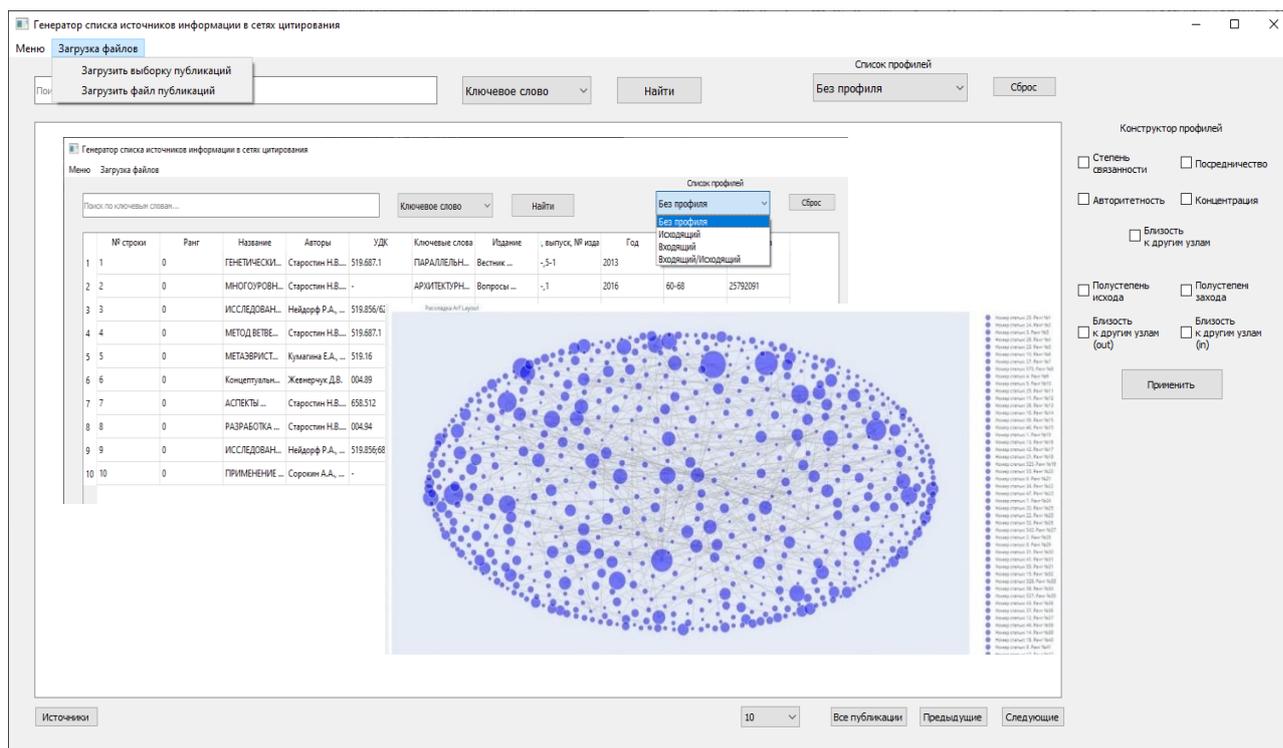


Рис. 2. Интерфейс программы «Генератор списка источников информации в сетях цитирования»

ВАРИАНТЫ ФОРМИРОВАНИЯ ПРОФИЛЯ ИССЛЕДОВАНИЯ

В зависимости от целей подбора и анализа публикаций в сети, используя различные комбинации параметров (мер центральностей), можно получить различные профили исследования h (формула (2)). При ранжировании публикаций по важности необходимо осуществить формирование профиля исследования, соответственно следует определить релевантные меры центральностей. Приведем примеры профилей исследования.

Профиль «Входящий/Исходящий». Включает в расчет показателя важности (C_{sum}) все меры центральностей, перечисленные в разделе «Отбор показателей для анализа сетей цитирования научных публикаций», без учета направленности графа. Выбор этого профиля позволяет отобрать публикации, которые могут быть как первоисточниками, так и обзорными, имеющими множество ссылок на другие публикации. Характеризует в целом уровень публикации, согласно предложенным критериям оценки с применением сетевого подхода.

Профиль «Исходящий». Включает в расчет показателя важности (C_{sum}) меры центральности, рассчитанные с учетом только исходящих связей, таких как полустепень исхода, центральность по близости к другим узлам в режиме *out*, центральность по информативности. Данный профиль может использоваться для отбора реферативных или обзорных публикаций.

Профиль «Входящий». Включает в расчет показателя важности (C_{sum}) меры центральности, рассчитанные с учетом только входящих связей, таких как полустепень захода, центральность по близости к другим узлам в режиме *in*, центральность по авторитетности. При выборе данного профиля представляется возможным отбор эмпирических публикаций, содержащих результаты оригинальных исследований.

Можно получить и другие профили, например, обратить в максимум только один показатель, а другие свести к минимуму, что характерно для любой сложной задачи исследования операций. В результате получим профиль «Авторитетность», ориентируясь только на меру центральности по авторитетности или профиль «Посредничество», учитывая меру центральности по посредничеству, которая управляет информацией среди других вершин графа сети через соединительный путь. При комбинировании двух показателей (меры центральности по близости к другим узлам и меры центральности по посредничеству) получим оценку важности узла относительно его положения в сети. Эти две меры центральности относят к геометрическим. Центральность по близости к другим узлам основана на кратчайших путях в графе и имеет простой физический смысл: чем меньше расстояния от исследуемой вершины до остальных вершин графа, тем больше будет значение самой центральности. Центральность по посредничеству позволяет достаточно хорошо определять «узкие места» в графе – вершины, входящие в состав ребра или набора ребер, соединяющих два ярко выраженных кластера.

РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

В настоящем разделе представлены полученные результаты анализа публикаций на основе данных международной базы цитирования Scopus. Сеть цитирования содержит статьи, изданные по 2023 год. Из данных о цитировании публикаций базы данных Scopus осуществлена выборка по ключевому слову

«Network Science». На основе полученных данных построен ориентированный граф, который содержит 850 вершин и 2169 дуг. На рисунке 3 представлен этот граф, для его визуализации использована программа VOSviewer.

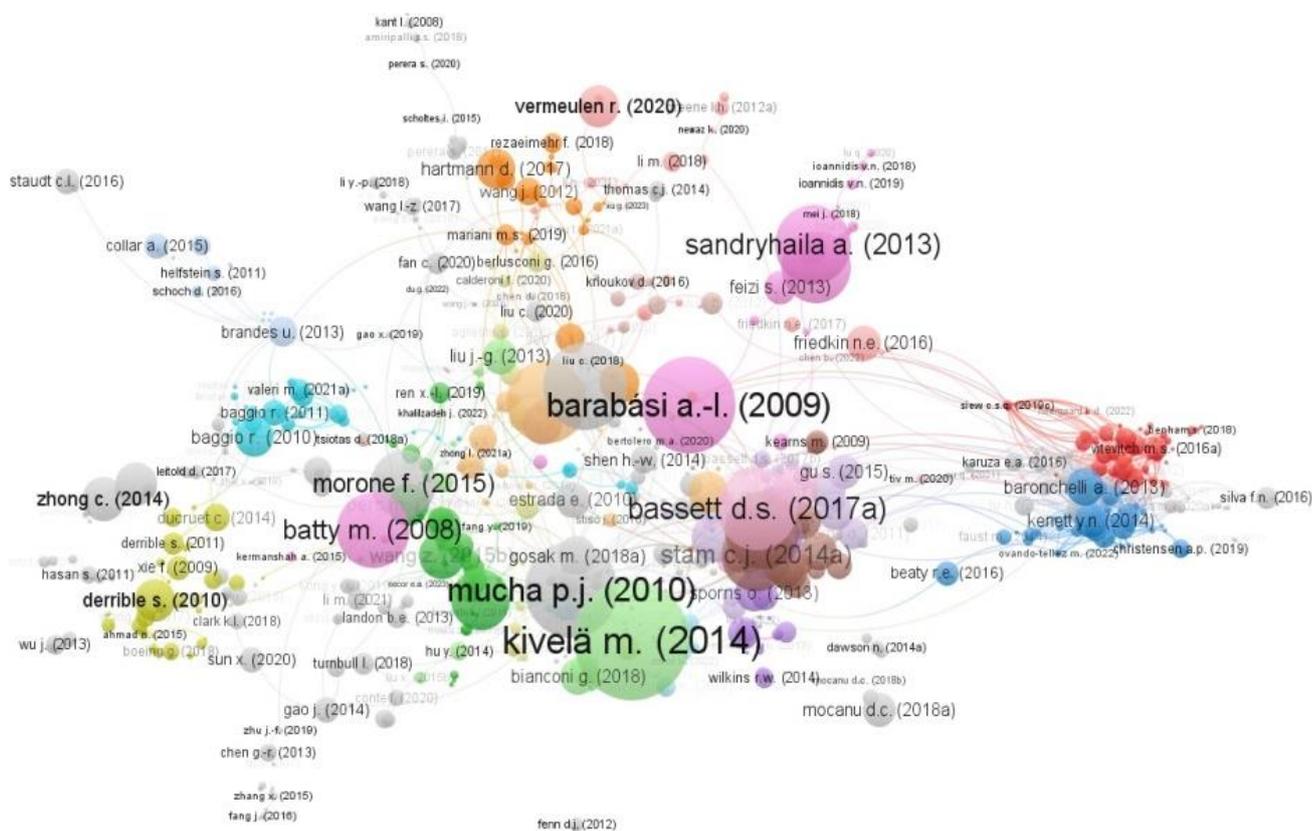


Рис. 3. Сеть цитирования по ключевому слову «Network Science»

На рисунке 4 представлен укрупненный фрагмент кластера данной сети цитирования, на который стоит обратить внимание. Из предыдущего изображения видно, что этот фрагмент даже визуально можно выделить в отдельный кластер. Диаметр узла сети пропорционален показателю количества цитирований публикации на рисунке 4 (входящие связи). На данном кластере произведем подробный анализ публикаций с помощью комбинирования показателей центральности узлов сети.

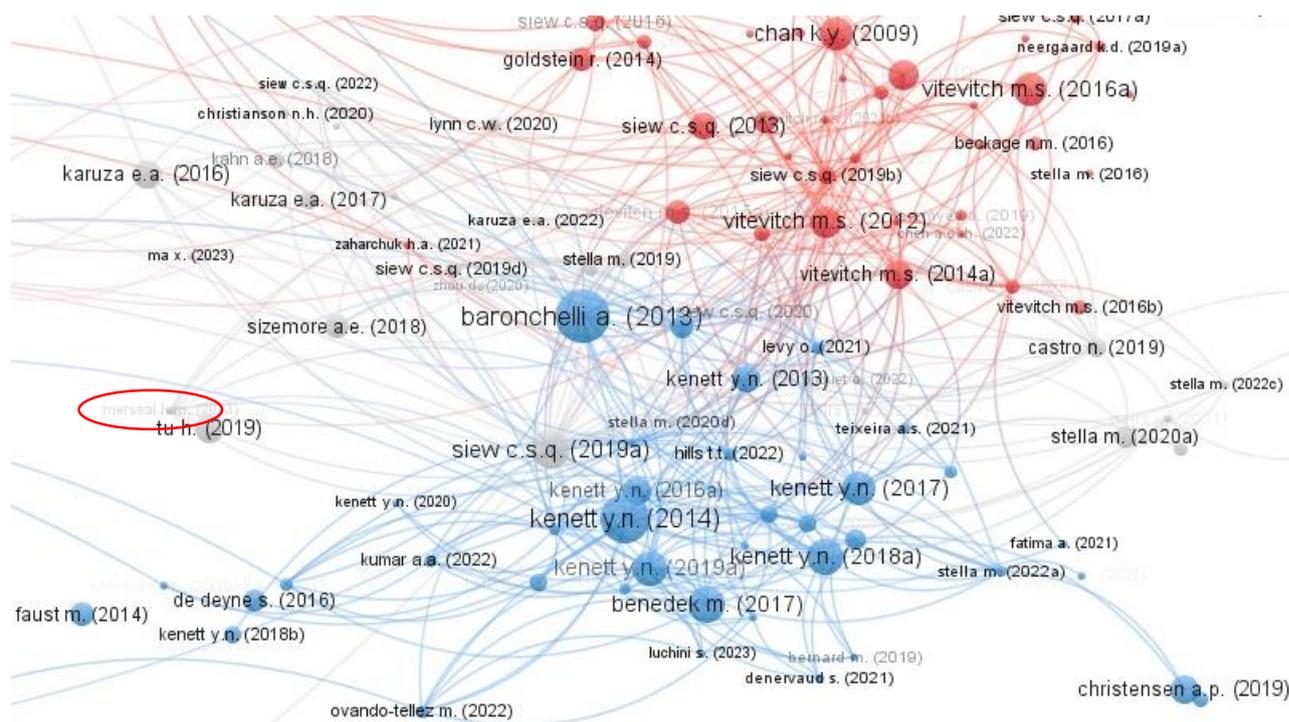


Рис. 4. Фрагмент сети цитирования по ключевому слову «Network Science»

Стоит обратить внимание на статью 2023 года автора Merseal H.M. (на рисунке 3 данный узел обведен красным маркером). Если использовать только возможности программы VOSviewer для исследования и визуализации сетей цитирования, то эту публикацию можно упустить из виду.

Выделенная публикация еще не имеет большого количества цитирований в связи с тем, что она только что опубликована. Однако она имеет важное значение в изучении и развитии науки о сетях. Тема данной статьи: «Представление мелодических взаимосвязей с использованием сетевой науки». Исходя из содержания публикаций, на которые она ссылается, можно судить о широком применении сетевого анализа и развитии науки о сетях (рисунок 5). Она цитирует авторитетные статьи, получившие большое количество цитирований, по следующим направлениям исследования:

- как сетевая наука может пролить свет на наше понимание познания;
- карты, транспортные средства и скайхуки в когнитивной сетевой науке;
- что сетевая наука может рассказать нам о фонологии и обработке языка;
- вклад современной сетевой науки в когнитивные науки;

- использование сетевой науки для понимания лексикона старения: связывание индивидуального опыта, семантических сетей и когнитивных способностей;
- значение статистического обучения для когнитивной сетевой науки.

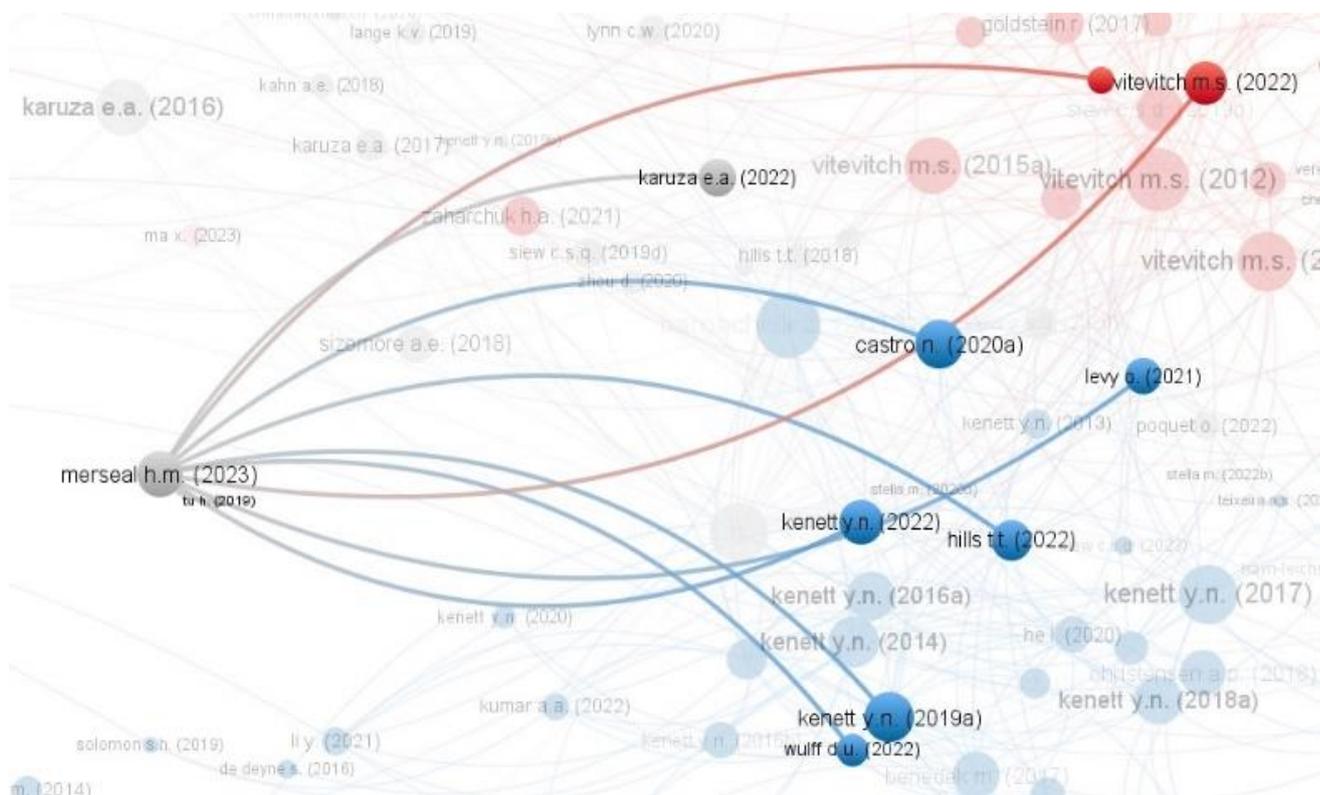


Рис. 5. Ссылки автора Н.М. Merseal на публикации

При изучении данной сети цитирования с помощью разработанной нами технологии, лежащей в основе программного комплекса анализа сетей цитирования, подобные статьи легко обнаружить. При использовании профиля исследования «Входящий/Исходящий» данная публикация получила 3 ранг из 390 (таблица 1).

В таблице 1 публикации проранжированы по убыванию их важности. Результат визуализирован с помощью разработанной программы визуализации графа цитирования и представлен на рисунке 6. Рассматриваемая публикация выделена красным маркером. Программа позволяет масштабировать размер вершины с учетом ранга публикации, полученного в результате комплексной оценки, с использованием математической модели (формула (2)). Радиус вер-

шины вычисляется через функцию натурального логарифма, что позволяет делать самые значимые вершины визуально различимыми. На данном рисунке показаны только те узлы, которые имеют ранг ≥ 50 , остальные вершины скрыты. Для демонстрации отображены подписи узлов с 1 по 3 ранг.

Таблица 1. Ранжированный список публикаций согласно профилю исследования «Входящий/Исходящий»

№ строки	Ранг	Название
468	1	Baronchelli A. (2013)
563	2	Karuza E.A. (2016)
369	3	Merseal H.M. (2023)
239	4	Kenett Y.N. (2022)
590	5	Vitevitch M.S. (2015)
466	6	Medaglia J.D. (2015)
440	7	Bassett D.S. (2017)
446	8	Stam C.J. (2014)
438	9	Mucha P.J. (2010)
535	10	Vitevitch M.S. (2012)
...
702	390	Leitold D. (2018)

С помощью данной программы визуализировать граф и осуществить масштабирование вершины графа согласно занимаемому публикацией рангу возможно на основе любого выбранного профиля исследования.

В таблице 2 представлены результаты ранжирования публикаций согласно нескольким профилям исследования. Для удобства описания название вершины заменено обозначением – S_i . С целью сравнения и анализа взяты первые три публикации из таблицы 1, занимающие 1, 2 и 3 ранги соответственно по профилю «Входящий/Исходящий» [28–30], и публикация [31] известного ученого по сетевой науке – Network Science, автора книги «Наука о сетях» (Cambridge, 2016).

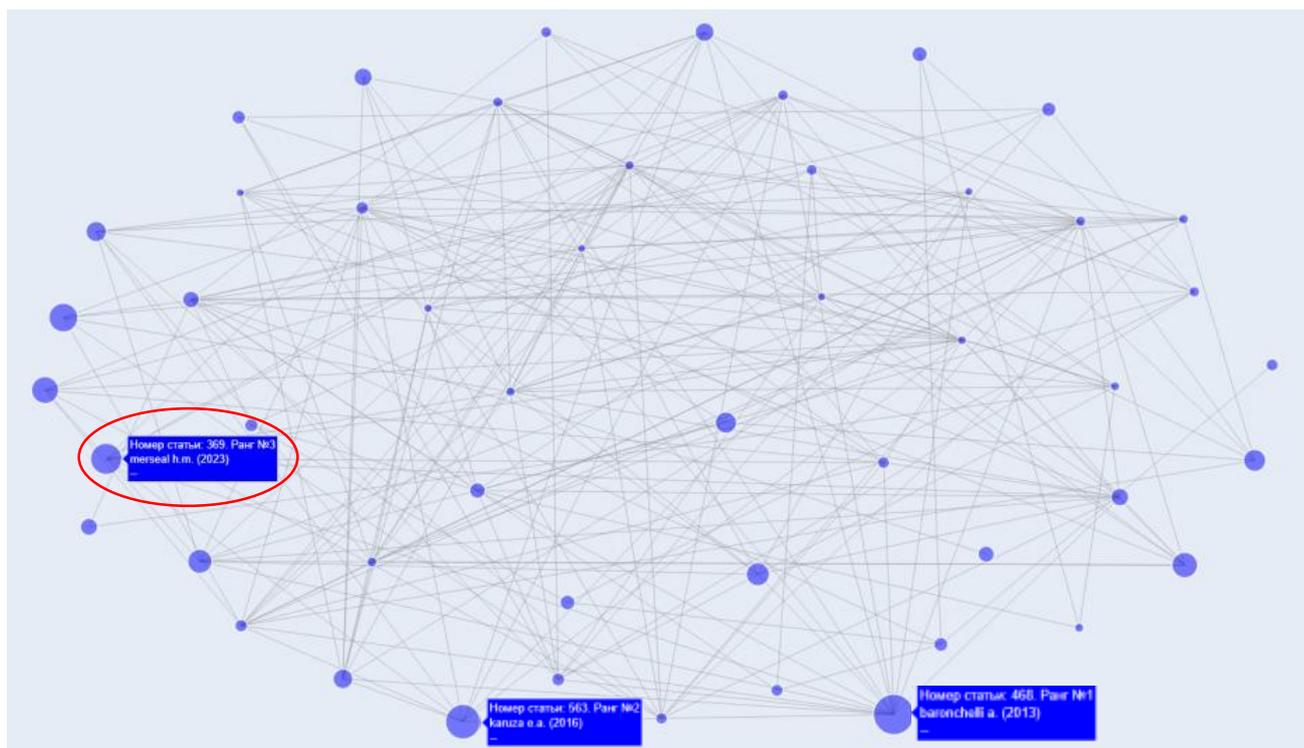


Рис. 6. Масштабирование вершины графа с учетом комплексной оценки (C_{sum}) центральности узлов сети согласно профилю «Входящий/Исходящий»

Таблица 2. Ранги публикаций по профилям исследования

Профиль исследования	Ранг вершин графа			
	Baronchelli A. (2013), S_1	Karuza E.A. (2016), S_2	Merseal H.M. (2023), S_3	Barabasi A.-L. (2009), S_4
Входящий/ Исходящий	1	2	3	120
Входящий	46	91	81	7
Исходящий	98	163	82	38
Авторитетность	4	7	37	76
Информативность	6	35	16	93
Степень связности	4	9	21	35
Полустепень захода	3	6	19	25
Полустепень исхода	7	14	15	23
Посредничество	17	3	12	137
Близости к другим узлам	85	78	56	393

Публикация S_4 занимает высокий ранг по профилю «Входящий», что говорит о важности данной работы как первоисточника информации о принятии теории сетей в качестве общей парадигмы, развитии науки о сетях как новой области исследований с особым набором задач и достижений. Ссылки на Albert-Laszlo Barabasi как авторитетного ученого имеются во всех трех публикациях, представленных в таблице. По сравнению с рассматриваемыми публикациями S_3 имеет наивысший ранг по профилю «Близость к другим узлам», что может характеризовать важность статьи относительно ее положения в исследуемом фрагменте графа. Количество ссылок в публикации S_1 составляет 174 источника, S_2 – 108 источников, S_3 – 101 источник, S_4 – всего 11. Публикация S_3 ссылается на S_1 , S_2 и S_3 и занимает достаточно высокий ранг по профилю «Информативность». Однако важность публикации по информативности у публикации S_3 значительно выше, чем у S_2 , это говорит о том, что S_3 ссылается на большее количество авторитетных статей. Значение меры центральности по информативности зависит от количества ребер, исходящих из исследуемой вершины, соединяющих вершины с высоким значением меры по авторитетности, и учитывает глубину связей, т. е. насколько эти вершины являются информативными. Публикация S_1 имеет высокий ранг по профилю «Авторитетность», что свидетельствует о том, что данная работа цитируется авторитетными авторами. Стоит отметить, что показатели рассчитаны, исходя из данных собранного фрагмента сети цитирования. В данном случае ограничивается данными о цитировании, полученными из базы данных Scopus.

Следовательно, можно сделать следующий вывод: исследуя топологию сети и применяя комбинированный метод анализа сетей цитирования, предложенный в данной работе, можно определить уровень научных публикаций в разных аспектах анализа, что дает представление о вкладе публикаций в исследуемой области науки.

ЗАКЛЮЧЕНИЕ

Выполнено систематическое исследование используемых в современной практике способов анализа важности научных публикаций. Отметим, что применение инструментов анализа научных публикаций на основе топологических свойств сети дает возможность осуществить их эффективный анализ. Для проведения экспериментальных исследований построена модель реальной сети цитирования публикаций. В связи с тем, что для любой прикладной задачи нахождения оптимального решения определяется свой набор критериев, разработана методика сетевого анализа научных публикаций на основе комплексной оценки узлов сети с учетом релевантных мер центральности, исходя из целей исследования. Внедрение подобных инструментов в наукометрические базы данных даст возможность осуществлять оценку публикаций, определяя их важность, не только на основе библиометрических показателей, в определении которых учитывается лишь количество цитирований. Ведь только библиометрические данные не могут служить критерием эффективности исследований или ценности публикаций.

Использование предложенных методов и алгоритмов анализа значимости научных публикаций позволяет повысить эффективность решения множества прикладных задач. Появилась возможность очень быстро проводить библиографические исследования для составления списка наиболее значимых литературных источников по какой-либо сфере научных исследований, определенной предметной области, по конкретной учебной дисциплине, составления литературного обзора и др. Использование предложенной методики применимо для: оптимизации комплектования литературой научных библиотек; подбора различных видов документов, в том числе патентов и патентных ландшафтов (объектов интеллектуальной собственности); поиска потенциальных соавторов или научных сообществ в исследуемой области знаний; наукометрии и инфометрики; управления научной и инновационной политикой.

Благодарность

Автор выражает благодарность Евгению Борисовичу Юдину, к. т. н., доценту кафедры «Математические методы и информационные технологии в экономике», начальнику Управления научной информацией ОмГТУ, за важнейшие советы при проведении исследования и оформлении статьи.

СПИСОК ЛИТЕРАТУРЫ

1. *Гонашвили А.С.* Наукометрические базы данных и работа с ними : науч.-метод. пособие / ун-т при Межпарламент. ассамблее ЕврАзЭС. СПб.: Изд-во ун-та при МПА ЕврАзЭС, 2020. 57 с.
2. H-index. URL: <https://en.wikipedia.org/wiki/H-index> (дата обращения: 15.04.2023).
3. Impact factor. URL: https://ru.wikipedia.org/wiki/Impact_factor (дата обращения: 15.04.2023).
4. *Lu Z., Ma Y., Song L.* Patent Citation Network Analysis Based on Improved Main Path Analysis: Mapping Key Technology Trajectory // *Advances in Artificial Intelligence and Security (ICAIS 2021): Communications in Computer and Information Science*. Springer: Cham, 2021. Vol. 1423. P. 158–171. https://doi.org/10.1007/978-3-030-78618-2_13.
5. *Wang J., Cheng Q., Lu W. et al.* A term function-aware keyword citation network method for science mapping analysis // *Information Processing & Management*. Vol. 60, no. 4. P. 103405. <https://doi.org/10.1016/j.ipm.2023.103405>.
6. *Ольгина И.Г., Пронин И.В., Абдрахманов А.Н.* Построение графовых моделей сети цитирования научных публикаций // *Системы управления, информационные технологии и математическое моделирование: материалы II Всерос. науч.-практ. конф. с междунар. участием (Омск, 19–20 мая 2020 г.)*. Омск: ОмГТУ, 2020. Т. I. С. 118–125.
7. *Zhao F., Zhang Y., Lu J. et al.* Measuring academic influence using heterogeneous author-citation networks // *Scientometrics*. 2019. Vol. 118. P. 1119–1140. <https://doi.org/10.1016/10.1007/s11192-019-03010-5>.
8. *Ji P., Jin J., Ke Z. T., L. W.* Co-citation and Co-authorship Networks of Statisticians // *Journal of Business & Economic Statistics*. 2022. Vol. 40, no. 2. P. 469–485.

<https://doi.org/10.1080/07350015.2021.1978469>.

9. *Luc P.T., Lan P.X., Le A.N.H., Tran B.T.* A Co-Citation and Co-Word Analysis of Social Entrepreneurship Research // *Journal of Social Entrepreneurship*. 2022. Vol. 13, No. 3. P. 324–339. <https://doi.org/10.1080/19420676.2020.1782971>.

10. *Печников А.А., Чебуков Д.Е.* Анализ соавторства в математических журналах Math-Net.Ru // *Научный сервис в сети Интернет: тр. XXIV Всерос. науч. конф. (19-22 сент. 2022 г.)*. М.: ИПМ им. М.В. Келдыша, 2022. С. 190-202. <https://doi.org/10.20948/abrau-2022-5>.

11. *Gómez S.* Centrality in Networks: Finding the Most Important Nodes // *Business and Consumer Analytics: New Ideas* / P. Moscato, N. Jane de Vries. Springer: Cham, 2019. P. 401–433. https://doi.org/10.1007/978-3-030-06222-4_8.

12. *Das K., Samanta S., Pa M.* Study on centrality measures in social networks: a survey // *Social Network Analysis and Mining*. 2018. Vol. 8. P. 13. <https://doi.org/10.1007/s13278-018-0493-2>.

13. *Бредихин С.В., Ляпунов В.М., Щербакова Н.Г.* Мера важности научной периодики – «центральность по посредничеству» // *Проблемы информатики*. 2014. №3. С. 53–64.

14. *Печников А.А., Чебуков Д.Е.* Структура графа цитирования журналов Math-Net.Ru // *Научный сервис в сети Интернет: тр. XXIII Всерос. науч. конф. (20–23 сент. 2021 г.)*. М.: ИПМ им. М.В. Келдыша, 2021. С. 265–278. <https://doi.org/10.20948/abrau-2021-2>.

15. *Ольгина И.Г.* Метод определения важных узлов сети цитирования научных публикаций // *Вестник компьютерных и информационных технологий*. 2021. Т. 18, № 5 (203). С. 3–10. <https://doi.org/10.14489/vkit.2021.05.pp.003-010>.

16. *Freeman L.C.* Centrality in social networks conceptual clarification // *Social Networks*. 1978. No. 31. P. 215–239.

17. *Newman M.E.J.* Scientific collaboration networks. I. Network construction and fundamental results // *Physical Review*. 2001. Vol. 64, No. 1. P. 016131. <https://doi.org/10.1103/PhysRevE.64.016131>.

18. *Brandes U.* A faster algorithm for betweenness centrality // *The Journal of Mathematical Sociology*. 2001. Vol. 25, No. 2. P. 163–177.

19. *Kleinberg J.* Authoritative sources in a hyperlinked environment // Proceedings of the 9th ACM-SIAM Symposium on Discrete Algorithms (SODA 98). 1998. P. 668–677.

20. *Leon C., Perez J.* Authority Centrality and Hub Centrality as metrics of systemic importance of financial market infrastructures // *Borradores de Economía*. 2013. Vol. 754. P. 1–25. <https://doi.org/10.2139/ssrn.2290271>.

21. *Бредихин С.В., Ляпунов В.М., Щербакова Н.Г., Юргенсон А.Н.* Параметры «центральности» узлов сети цитирования научных статей // *Проблемы информатики*. 2016. № 1. С. 39–57.

22. *Shibata N., Kajikawa Y., Takeda Y. et al.* Early detection of innovations from citation networks // International Conference on Industrial Engineering and Engineering Management (Hong Kong, 8–11 December 2009). IEEE, 2009. <https://doi.org/10.1109/ieem.2009.5373444>.

23. *Baglioni M., Geraci F., Pellegrini M., Lastres E.* Fast Exact Computation of betweenness Centrality in Social Networks // *Advances in Social Networks Analysis and Mining: International Conference 2012 IEEE/ACM (Istanbul, 26–29 August 2012)*. P. 450–456. <https://doi.org/10.1109/ASONAM.2012.79>.

24. *Farhan M.T., Darwiyanto E., Asror I.* Analysis of Hubs and Authorities Centrality Using Probabilistic Affinity Index (PAI) on directed-weighted graph in Social Network Analysis // *Journal of Physics: Conference Series*. 2019. Vol. 1192. P. 012005. <https://doi.org/10.1088/1742-6596/1192/1/012005>.

25. *Marra A., Antonelli P., Dell’Anna L., Pozzi C.* A network analysis using metadata to investigate innovation in clean-tech – Implications for energy policy // *Energy Policy*. 2015. Vol. 86. P. 17–26.

26. *Baronchelli A., Ferrer-i-Cancho R., Pastor-Satorras R., Chater N., Christiansen Morten N.* Networks in cognitive science // *Trends in cognitive sciences*. 2013. Vol. 17. Iss. 7. P. 348–360. <https://doi.org/10.1016/j.tics.2013.04.010>.

27. *Karuza E.A., Thompson-Schill Sh.L., Bassett Danielle S.* Local patterns to global architectures: influences of network topology on human learning // *Trends in cognitive sciences*. 2016. Vol. 20. Iss. 8. P. 629–640. <https://doi.org/10.1016/j.tics.2016.06.003>.

28. Merseal Hannah M., Beaty Roger E., Kenett Yoed N., Lloyd-Co J., Orjan de Manzano, Norgaard Martin. Representing melodic relationships using network science // *Cognition*. 2023. Vol. 233. P. 105362.

<https://doi.org/10.1016/j.cognition.2022.105362>.

29. Barabasi A.-L. Scale-free networks: Aa decade and beyond // *Science*. 2009. Vol. 325. Iss. 5939. P. 412–413.

30. Кошелева Н.Н. Корреляционный анализ и его применение для подсчета ранговой корреляции Спирмена // *Актуальные проблемы гуманитарных и естественных наук*. 2012. № 5. С. 23–26.

31. Ермолаев О.Ю. Математическая статистика для психологов. М.: Московский психолого-социальный институт: Флинта, 2003. 366 с.

METHODOLOGY OF NETWORK ANALYSIS OF SCIENTIFIC PUBLICATIONS

I. G. Olgina^[0000-0002-9932-4552]

Omsk State Technical University

inna_olgina@mail.ru

Abstract

The relevance of the issues of the analysis of scientific publications is due to the fact that with the advent of Internet technologies, it became possible to collect data on the publication citation network. Meanwhile, the current approach to the analysis of scientific publications is based on bibliometric indicators that take into account only the number of citations. However, network analysis, which is mainly used in the study of social networks, is becoming increasingly widely used. The author has developed a methodology that allows for an effective analysis of scientific publications based on network analysis methods alternative to bibliometric methods. As criteria for evaluating scientific publications based on network analysis, relevant measures of the centrality of the citation network nodes are established: centrality by degree of connectivity; centrality by proximity to other nodes; centrality by mediation; centrality by authority; centrality by concentration. The author presented the experiment re-

sult that allows validating the developed methodology of network analysis of the scientific publications significance. Scientometric databases were used as primary sources of data on publications, which make it possible to track the citation of publications and identify relevant citation networks. The application of the proposed network analysis methodology contributes to the identification of important publications in the development of the scientific direction.

Keywords: *citation network, publications, scientometrics, bibliometric analysis, network analysis, graph*

REFERENCES

1. Gonashvili A.S. Naukometricheskie bazy dannyh i rabota s nimi [Scientometric databases and working with them]. SPb.: MPA EvrAzJeS. 2020. 57 p. (In Russ.).

2. H-index. Available at: <https://en.wikipedia.org/wiki/H-index> (accessed 15.04.2023).

3. Impact factor. Available at: https://ru.wikipedia.org/wiki/Impact_factor (accessed 15.04.2023).

4. Lu Z., Ma Y., Song L. Patent Citation Network Analysis Based on Improved Main Path Analysis: Mapping Key Technology Trajectory // *Advances in Artificial Intelligence and Security (ICAIS 2021): Communications in Computer and Information Science*. Springer: Cham, 2021. Vol. 1423. P. 158–171. https://doi.org/10.1007/978-3-030-78618-2_13.

5. Wang J., Cheng Q., Lu W. et al. A term function-aware keyword citation network method for science mapping analysis // *Information Processing & Management*. Vol. 60, No. 4. P. 103405. <https://doi.org/10.1016/j.ipm.2023.103405>.

6. Ol'gina I.G., Pronin I.V., Abdrahmanov A.N. Postroenie grafovyyh modelej seti citirovaniya nauchnyh publikacij [Graph models construction of the citation network of scientific publications] // *Sistemy upravleniya, informacionnye tehnologii i matematicheskoe modelirovanie: materialy Vtoroj Vseros. nauch.-prakt. konf. s mezhdunar. uchast* [Control systems, information technologies and mathematical modeling: Collected papers]. Omsk: OmGTU, 2020. Vol. 1. P. 118–125 (In Russ.).

7. Zhao F., Zhang Y., Lu J. et al. Measuring academic influence using heterogeneous author-citation networks // *Scientometrics*. 2019. Vol. 118. P. 1119–1140. <https://doi.org/10.1016/10.1007/s11192-019-03010-5>.

8. Ji P., Jin J., Ke Z. T., L. W. Co-citation and Co-authorship Networks of Statisticians // Journal of Business & Economic Statistics. 2022. Vol. 40, No. 2. P. 469–485. <https://doi.org/10.1080/07350015.2021.1978469>.

9. Luc P.T., Lan P.X., Le A.N.H., Tran B.T. A Co-Citation and Co-Word Analysis of Social Entrepreneurship Research // Journal of Social Entrepreneurship. 2022. Vol. 13, No. 3. P. 324–339. <https://doi.org/10.1080/19420676.2020.1782971>.

10. Pechnikov A.A., Chebukov D.E. Analiz soavtorstva v matematicheskikh zhurnalakh Math-Net.Ru [Analysis of co-authorship in mathematical journals of Math-Net.Ru]. Nauchnyj servis v seti Internet: trudy dvadcat' tret'ej Vserossijskoj nauchnoj konferencii. [Scientific service on the Internet: Conference proceedings]. M., 2022. P. 190-202. <https://doi.org/10.20948/abrau-2022-5>.

11. Gómez S. Centrality in Networks: Finding the Most Important Nodes // Business and Consumer Analytics: New Ideas / P. Moscato, N. Jane de Vries. Springer: Cham, 2019. P. 401–433. https://doi.org/10.1007/978-3-030-06222-4_8.

12. Das K., Samanta S., Pa M. Study on centrality measures in social networks: a survey // Social Network Analysis and Mining. 2018. Vol. 8. P. 13. <https://doi.org/10.1007/s13278-018-0493-2>.

13. Bredihin S.V., Ljapunov V.M., Shherbakova N.G. Mera vazhnosti nauchnoj periodiki – central'nost' po posrednichestvu [Measurement of the scientific periodicals importance: cooperation centrality]. Problemy informatiki = Problems of Informatics. 2014. No. 3. P. 53–64.

14. Pechnikov A.A., Chebukov D.E. Struktura grafa citirovanija zhurnalov Math-Net.Ru [Structure of the citation graph of Math-Net.Ru journals]. Nauchnyj servis v seti Internet: trudy dvadcat' tret'ej Vserossijskoj nauchnoj konferencii. [Scientific service on the Internet: Conference proceedings]. M., 2021. P. 265–278. <https://doi.org/10.20948/abrau-2021-2> (In Russ.).

15. Ol'gina I.G. Metod opredelenija vazhnyh uzlov seti citirovanija nauchnyh publikacij [Method for determining important nodes of the citation network of scientific publications]. Vestnik komp'juternyh i informacionnyh tehnologij = Herald of Computer and Information Technologies. 2021. Vol. 18, No. 5 (203). P. 3–10. <https://doi.org/10.14489/vkit.2021.05.pp.003-010> (In Russ.).

16. *Freeman L.C.* Centrality in social networks conceptual clarification // *Social Networks*. 1978. No. 31. P. 215–239.

17. *Newman M.E.J.* Scientific collaboration networks. I. Network construction and fundamental results // *Physical Review*. 2001. Vol. 64, No. 1. P. 016131. <https://doi.org/10.1103/PhysRevE.64.016131>.

18. *Brandes U.* A faster algorithm for betweenness centrality // *The Journal of Mathematical Sociology*. 2001. Vol. 25, No. 2. P. 163–177.

19. *Kleinberg J.* Authoritative sources in a hyperlinked environment // *Proceedings of the 9th ACM-SIAM Symposium on Discrete Algorithms (SODA 98)*. 1998. P. 668–677.

20. *Leon C., Perez J.* Authority Centrality and Hub Centrality as metrics of systemic importance of financial market infrastructures // *Borradores de Economía*. 2013. Vol. 754. P. 1–25. <https://doi.org/10.2139/ssrn.2290271>.

21. Бредихин С.В., Ляпунов В.М., Щербакова Н.Г., Юргенсон А.Н. Параметры «центральности» узлов сети цитирования научных статей // *Проблемы информатики*. 2016. № 1. С. 39–57.

22. *Shibata N., Kajikawa Y., Takeda Y. et al.* Early detection of innovations from citation networks // *International Conference on Industrial Engineering and Engineering Management (Hong Kong, 8–11 December 2009)*. IEEE, 2009. <https://doi.org/10.1109/ieem.2009.5373444>.

23. *Baglioni M., Geraci F., Pellegrini M., Lastres E.* Fast Exact Computation of betweenness Centrality in Social Networks // *Advances in Social Networks Analysis and Mining: International Conference 2012 IEEE/ACM (Istanbul, 26–29 August 2012)*. P. 450–456. <https://doi.org/10.1109/ASONAM.2012.79>.

24. *Farhan M.T., Darwiyanto E., Asror I.* Analysis of Hubs and Authorities Centrality Using Probabilistic Affinity Index (PAI) on directed-weighted graph in Social Network Analysis // *Journal of Physics: Conference Series*. 2019. Vol. 1192. P. 012005. <https://doi.org/10.1088/1742-6596/1192/1/012005>.

25. *Marra A., Antonelli P., Dell'Anna L., Pozzi C.* A network analysis using metadata to investigate innovation in clean-tech – Implications for energy policy // *Energy Policy*. 2015. Vol. 86. P. 17–26.

26. *Kosheleva N.N.* Korreljacionnyj analiz i ego primenenie dlja podscheta rangovoj korreljicii Spirmena [Correlation analysis and its application for calculate Spearman's rank correlation] // Aktual'nye problemy gumanitarnyh i estestvennyh nauk = Current problems in the humanities and natural sciences. 2012. No. 5. P. 23–26. (In Russ.).

27. *Ermolaev O.Ju.* Matematicheskaja statistika dlja psihologov [Mathematical statistics for psychologists]. M.: Moscow Psychological and Social Institute: Flinta, 2003. 366 p. (In Russ.).

28. *Baronchelli A., Ferrer-i-Cancho R., Pastor-Satorras R., Chater N., Christiansen Morten N.* Networks in cognitive science // Trends in cognitive sciences. 2013. Vol. 17. Iss. 7. P. 348–360. <https://doi.org/10.1016/j.tics.2013.04.010>.

29. *Karuza E.A., Thompson-Schill Sh.L., Bassett Danielle S.* Local patterns to global architectures: influences of network topology on human learning // Trends in cognitive sciences. 2016. Vol. 20. Iss. 8. P. 629–640. <https://doi.org/10.1016/j.tics.2016.06.003>.

30. *Merseal Hannah M., Beaty Roger E., Kenett Yoed N., Lloyd-Co J., Orjan de Manzano, Norgaard Martin.* Representing melodic relationships using network science // Cognition. 2023. Vol. 233. P. 105362.

<https://doi.org/10.1016/j.cognition.2022.105362>.

31. *Barabasi A.-L.* Scale-free networks: Aa decade and beyond // Science. 2009. Vol. 325. Iss. 5939. P. 412–413.

СВЕДЕНИЯ ОБ АВТОРЕ



ОЛЬГИНА Инна Геннадьевна – директор библиотеки ОмГТУ, старший преподаватель кафедры «Математические методы и информационные технологии в экономике», ОмГТУ. Область научных интересов: теория графов, применение методов системного анализа для исследования сетей цитирования.

Inna Gennadevna OLGINA – Director of the library in Omsk State Technical University, Senior Lecturer of the Mathematical Methods and Information Technologies in Economics Department, Omsk State Technical University. Research interests: graph theory, application of systems analysis methods for the citation networks research.

email: inna_olgina@mail.ru

ORCID: 0000-0002-9932-4552

Материал поступил в редакцию 20 августа 2023 года