

СИСТЕМА ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПРИ ВЫБОРЕ ИСТОЧНИКОВ ИНФОРМАЦИИ В СЕТЯХ ЦИТИРОВАНИЯ

И. Г. Ольгина^[0000-0002-9932-4552]

Омский государственный технический университет

inna_olgina@mail.ru

Аннотация

С появлением науки о сетях стало возможным исследовать сложные сетевые системы, в том числе социальные и информационные, посредством представления их в виде графовых моделей. Рост в геометрической прогрессии общего объема научных публикаций обуславливает актуальность задач анализа их взаимосвязей. В науке о сетях для решения данных задач разрабатываются модели и методы, относящиеся к сфере так называемых сетей цитирования. Однако сетевые метрики не используются при анализе публикаций в базах цитирования.

В работе рассмотрены вопросы создания системы поддержки принятия решений при выборе источников информации на основе данных о цитировании научных публикаций. Разработан программный комплекс для принятия решений по определению важности публикации в определенной тематической области. В основу работы этого программного комплекса заложен метод ранжирования публикаций по важности на основе анализа сетей цитирования, позволяющий выявить публикации, которые явно не выделяются в чистом виде при ранжировании на основе известных библиометрических показателей или известных мер центральности узлов. Проведены исследование и сравнительный анализ программного обеспечения для визуализации и исследования всех видов графов и социальных сетей. Выполнены исследования, подтверждающие эффективность предлагаемой системы поддержки принятия решений при выборе источников информации.

Ключевые слова: сеть цитирования, публикация, наукометрия, система поддержки принятия решений, архитектура программного комплекса, сетевой анализ, граф.

ВВЕДЕНИЕ

Теория сетей имеет приложения во многих дисциплинах. История становления и развития науки о сетях начинается с 1736 года [1]. В 1936 году Денеш Кёниг, венгерский профессор математики, написал первую книгу по теории графов, озаглавленную «Теория конечных и бесконечных графов» [2]. Широкое распространение начинает получать сетевой анализ, применяемый преимущественно в исследовании социальных сетей посредством представления их в виде графовых моделей.

При возрастающем количестве информации в виде научных публикаций возникает потребность в более глубоком их анализе с помощью сетевого подхода. Следует отметить, что в настоящее время отсутствует способ анализа публикаций, позволяющий учесть все необходимые структурно-топологические особенности сети, для решения таких задач в содержательной интерпретации, как определение обзорных статей, эмпирических и важных статей по определенной тематике.

Актуальность темы настоящего исследования обусловлена необходимостью разработки технологии анализа научных публикаций, основанной на исследовании показателей сетей цитирования (сетевых метриках), которые являются разновидностью социальных сетей.

Цель проведенного исследования – разработка системы поддержки принятия решений на основе анализа сетей цитирования публикаций для выбора источников информации, на которые в первую очередь стоит обратить внимание на этапе исследования предметной области. Под источниками информации подразумеваются научные публикации (статьи, патенты, материалы конференций, научные журналы и книги). Входными данными для названной системы являются публикации и выявленные взаимосвязи между ними – цитирования.

ОБЗОР ПРЕДМЕТНОЙ ОБЛАСТИ

Сеть цитирования представляет собой ориентированный граф, который связывает цитаты в коллекции документов. Каждая вершина (или узел) на графе представляет документ в коллекции, и каждое ребро направлено от одного документа к другому, на который он ссылается (или наоборот, в зависимости от конкретной реализации) [3].

В 2021 г. опубликована монография [4], в которой рассмотрены основные понятия сети и цитирования, построения моделей сети в виде неориентированных и ориентированных графов, представлена формализация понятия меры центральности. В 1977 г. L. Freeman [5] впервые предложил интуитивные концепции, положенные в основу определений мер центральности акторов. Из предложенных им определений следует, что центральность узла сети – мера его важности в сети. Характеристика «центральность» позволяет определить степень важности вершины графа, основываясь на ее расположении.

В 2021 году опубликована статья [6], авторы которой представили способ построения графа цитирования журналов с учетом весов дуг следующим образом: математической моделью сети цитирования журналов является ориентированный граф $G_{cit}(V, E, W)$, где V – множество вершин (соответствующих журналам), E – множество дуг (связывают пары вершин i и j , если статьи журнала i имеют хотя бы одну ссылку на статьи журнала j), W – множество весов дуг (вес $w(i, j)$ дуги $e(i, j)$ равен количеству ссылок, сделанных со всех статей журнала i на статьи журнала j).

Значимость научных журналов характеризуется посредством их ранжирования по значениям показателей, построенных на основе данных о цитировании статей. Для графа цитирования в [6] с целью ранжирования взят показатель центральности узла PageRank (PR) и осуществлено сравнение ранжирования журналов на основе данного показателя с рейтингом журналов SCIENCE INDEX с помощью метода ранговой корреляции Спирмена. Полученные результаты свидетельствуют об умеренной прямой связи между двумя ранжированиями.

В статьях [7–11] даны примеры анализа сетей цитирования статей. В работе [7] приведена содержательная интерпретация важности вершины по *hub centrality* (мере центральности по информативности). Это статьи, имеющие высокую степень *hub centrality* и являющиеся реферативными статьями и обзорами. Стоит

отметить, что алгоритм вычисления степени *hub centrality* работает в рамках определенной компоненты связности графа.

Широкое применение получил анализ сетей цитирования публикаций среди западных ученых. Исследователи из Китая и США в 2020 году в работе [12] предложили модификацию меры центральности PageRank. Для сравнения разработанных расширений метода PageRank авторы вычисляли коэффициент ранговой корреляции и выбирали метод с наилучшей корреляцией. Затем было проведено сравнение рейтингов по новым методам с реальным использованием (т. е. с посещениями веб-сайта GenBank и загрузками Figshare). Для генерации сети был разработан алгоритм сбора данных по цифровому идентификатору объекта (*digital object identifier (DOI)*), в результате осуществлялось построение сети цитирования DOI-DOI. Недостатком этого алгоритма является ограничение возможности формирования полной сети: в сеть не будут включены статьи, не имеющие DOI. Например, могут быть не охвачены публикации, изданные в ранние годы, или статьи, которым не присваивался DOI. Соответственно алгоритм не применим для российских баз цитирования.

Стоит отметить, что применение алгоритма PageRank для анализа научных публикаций рассмотрено во многих зарубежных статьях (например, [12–17]). PageRank решает проблему, связанную с тем, что сети цитирования всегда направлены назад во времени. В этой модели каждая вершина графа является публикацией. Цель этого метода – оценить влияние вершины через ее цитирования.

Существует специализированное программное обеспечение (ПО) для визуализации и исследования всех видов графов и социальных сетей. Цель его создания состоит в том, чтобы помочь аналитикам данных выдвигать гипотезы, интуитивно обнаруживать закономерности, выявлять структурные особенности или ошибки во время поиска данных. Это дополнительный инструмент к традиционной статистике, поскольку визуальное мышление с интерактивными интерфейсами признано удобным для облегчения рассуждений.

Сравнительный анализ ПО для визуализации и исследования всех видов графов и социальных сетей представлен в работе [18] и приведен в таблице 1. Отметим, что существующее ПО – только зарубежных производителей, поэтому

актуальна разработка российского продукта, адаптированного для анализа публикаций.

Таблица 1. Сравнительный анализ программного обеспечения для визуализации и исследования всех видов графов и социальных сетей

	Gephi	Sentinel visualizer	TomSawyer Perspectives	VOSviewer
Показатели	центральность по посредничеству, по близости к другим узлам			количество цитирований количество ссылок
	коэффициент кластеризации, PageRank, обнаружение сообщества, кратчайший путь	центральность по степени связности центральность по информативности, центральность по авторитетности, собственное значение		
Сильные стороны	Глубокая настройка внешнего вида сети. Работа алгоритмов обнаружения структурных шаблонов	Анализ социальных сетей, структуры данных, геопространства. Интеллектуальные алгоритмы для обнаружения аномалий, модели поведения	Обработка крупномасштабных графов, быстрое отображение и манипулирование данными, глубокая настройка внешнего вида сети	Работа с научными публикациями
Слабые стороны	Нет возможности комплексной оценки с учетом нескольких показателей			Невозможен анализ топологии сети
Стоимость	Бесплатно	Платно	Платно	Бесплатно
Страна	Франция	США	Америка	Нидерланды

ЭТАПЫ ПРОЕКТИРОВАНИЯ И АРХИТЕКТУРА СИСТЕМЫ АНАЛИЗА ИСТОЧНИКОВ ИНФОРМАЦИИ В СЕТЯХ ЦИТИРОВАНИЯ

Рассмотрим этапы проектирования системы поддержки принятия решений (СППР) при выборе источников информации, позволяющей автоматизировать обработку и анализ данных о публикациях и принять решение о важности источника информации в определенной тематической области (рисунок 2).



Рис. 2. Этапы разработки СППР при выборе источников информации

На рисунке 3 представлена архитектура программного комплекса СППР, реализованного с использованием языка программирования Python. Каждый программно-алгоритмический модуль комплекса содержит функциональные блоки, используемые на определенном этапе проектирования системы принятия решений при выборе источников информации.

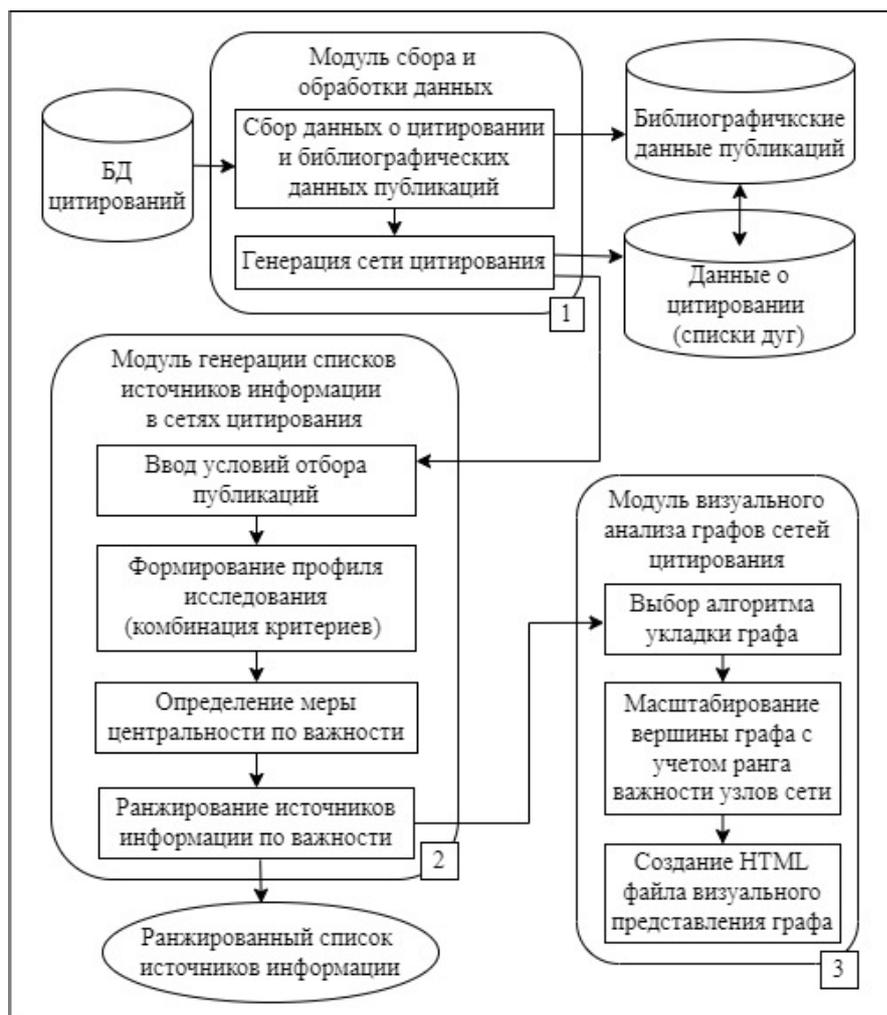


Рис. 3. Архитектура программного комплекса для разработки СППР при выборе источников информации в сетях цитирования

Рассмотрим процессы разработки СППР (рис. 2), реализуемые соответствующими программно-алгоритмическими модулями (рис. 3).

На *первом этапе* с помощью модуля (1) осуществляются сбор и обработка данных о цитировании публикаций. Программный модуль позволяет получать сеть цитирования в виде ориентированного графа. Реализована возможность остановки и продолжения сбора данных в любой момент (без потери данных), получения библиографических данных публикаций по номеру вершины графа. Реализована возможность записи библиографических данных публикаций и изданий, на которые ссылаются авторы, в таблице Microsoft Excel [19].

Затем на *втором этапе* обработанные данные о цитировании поступают на вход модуля (2) для формирования множества релевантных показателей для

выбора важной публикации согласно выбранной цели исследования. Далее в модуле (2) на *третьем этапе* осуществляется формирование профиля исследования, исходя из цели. На *четвертом этапе* осуществляется оценка источников информации, которая вычисляется методом ранжирования научных публикаций по важности [20].

Результаты работы модуля (2) проверялись на согласованность с результатами известных методов: метода ранговой корреляции Спирмена и Стьюдента [21] и метода анализа иерархий (МАИ) [22].

На *пятом этапе* в модуле (3) полученные результаты используются для просмотра и визуального анализа графа сети цитирования на основе исследования и ранжирования публикаций по важности. Этот модуль также может использоваться в базах данных научного цитирования, в различных типах и видах библиотек, вузах и научно-исследовательских институтах [23].

Интерфейс программного комплекса представлен на рисунке 4.

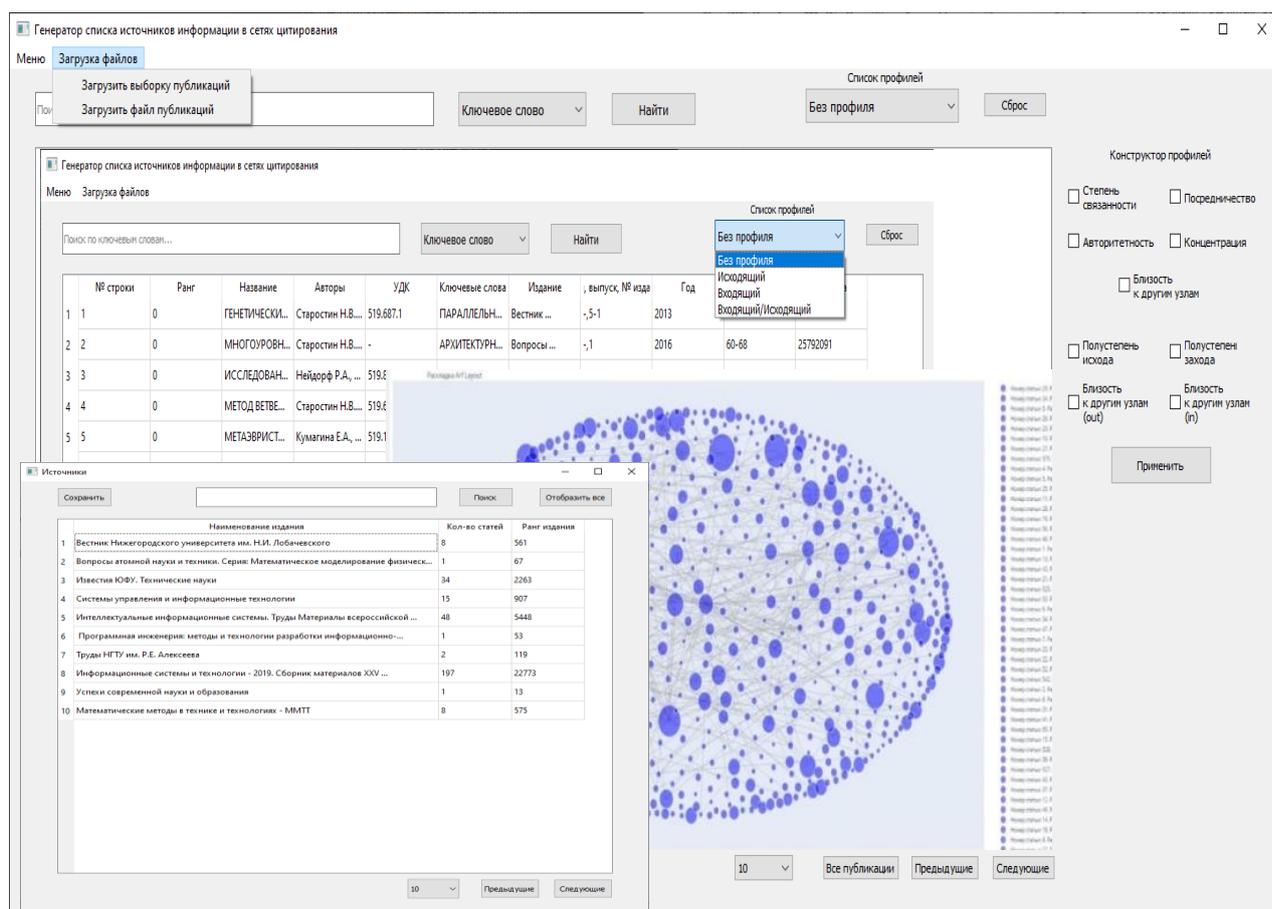


Рис. 4. Интерфейс программного комплекса СППР

РЕАЛИЗАЦИЯ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПРИ ВЫБОРЕ ИСТОЧНИКОВ ИНФОРМАЦИИ

Этап 1. Сбор данных о цитировании, создание графовых моделей и библиографических данных о публикации

Основной задачей выявления структурной информации о сети цитирования является сбор данных. С целью получения и обработки данных из различных баз данных цитирования разработан соответствующий модуль (1). В основу его работы заложен алгоритм сбора данных для построения сети цитирования публикаций. Этот алгоритм подробно описан в работе [24]. Разработанный алгоритм программы имеет преимущества при построении графовых моделей сети цитирования в виде списков дуг, которые записываются в файл с расширением .net. Используя данный алгоритм, можно получить сеть, в которую входят все источники информации, учитывая всю глубину связей. Кроме того, алгоритм позволяет получать сеть цитирования публикаций, включая все публикации по заданной тематике, еще не имеющие цитирований.

В первом блоке модуля (1) производится сбор данных. Во втором блоке модуля осуществляется генерация графа сети цитирования.

Использованы два способа представления графа сети цитирования: в виде матрицы смежности и в виде списка дуг. Исследованы и созданы условия для построения модели сети цитирования (рис. 5).

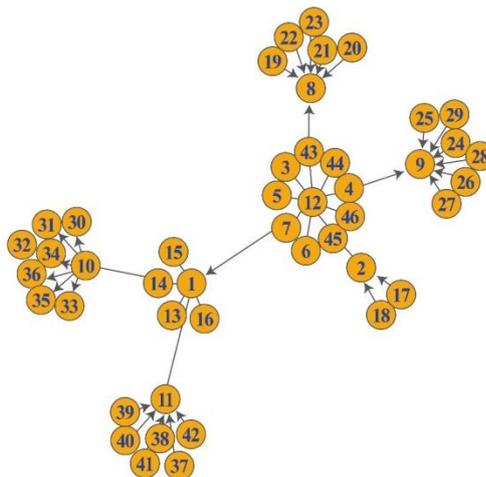


Рис. 5. Фрагмент сети цитирования научных публикаций библиографической базы данных RePEC

На рисунке 5 представлен пример графа сети цитирования, полученный на основе данных цитирования библиографической БД RePEc [25]. БД RePEc обеспечивает доступ к более 850 000 различным материалам по экономике, финансам, менеджменту и маркетингу. Из данных о цитировании публикаций БД RePEc выделена компонента связности, содержащая 46 узлов.

Этап 2. Формирование множества показателей для анализа сетей цитирования

Проведен анализ способов расчета мер центральности узлов сети, используемых в современной практике. Показано, что для выявления значимой роли узлов сети цитирования недостаточно одной из мер центральности, и на данный момент не существует универсальной меры.

Для определения значимой роли узла сети цитирования достаточно учитывать следующие меры центральности:

- центральность по степени связности (degree centrality);
- центральность по близости к другим узлам (closeness centrality);
- центральность по посредничеству (betweenness centrality);
- центральность по авторитетности (authority centrality);
- центральность по информативности (hub centrality).

В работе [7] рассмотрены эти меры центральности именно для ранжирования узлов сети цитирования научных публикаций. В связи с тем, что сеть цитирования является ориентированной, входящие и исходящие связи анализируются отдельно. При определении центральности по близости к другим узлам для орграфов можно рассматривать как дистанции от определяемой вершины до всех остальных (режим *out*), так и дистанции от всех вершин до определяемой (режим *in*). В случае, когда цитируемым объектам придается большая значимость, имеет смысл пользоваться вторым определением.

Этап 3. Формирование профиля исследования и расчет комплексного показателя важности публикации

Полученное множество показателей рассмотрено с целью формирования профиля исследования. *Профиль исследования* – это комплекс мер центральности узла сети цитирования с установленным индексом релевантности для каждой меры, имеющих общую задачу. Исходя из профиля исследования, можно

определить важные публикации, отвечающие поставленной задаче. Для этого создан программно-алгоритмический модуль (2).

Особенность данного модуля заключается в созданном конструкторе профилей, с помощью которого можно выбирать уже готовые профили исследования либо формировать собственные, исходя из целей отбора публикаций. На основе осуществленного отбора релевантных мер центральности узлов можно сформировать, например, следующие профили исследования, соответствующие таким целям, как: лучшие идеи (авторитетность и входящее цитирование), лучшие обзоры (исходящее цитирование и информативность), публикации-коммуникаторы (близость и промежуточность). Названный конструктор профилей реализован во *втором блоке* модуля (2).

В связи с тем, что граф сети цитирования может содержать большое количество вершин и дуг, для сокращения времени обработки и визуализации данных в модуле (2) в *первом блоке* предусмотрены параметры отбора: по тематике, ключевым словам, годам издания, авторам, аффилиациям организаций, типам и видам изданий.

В *третьем блоке* модуля (2) реализован расчет комплексного показателя важности публикации

$$C^h(v) = \sum_{i=1}^n k_i^h R_i(v), \quad (3)$$

где C^h – мера центральности по важности соответствующего профиля исследования; v – узел сети; R_i – ранг i -го показателя; n – число показателей; k_i^h – весовой коэффициент – индекс релевантности i -го показателя, $0 \leq k_i \leq 1$, $k_i^h \in Q$; h – профиль исследования.

Разработанная математическая модель определения важности узлов сети цитирования позволяет учитывать все структурно-топологические характеристики заданного фрагмента сети в комплексе и определять комплексную оценку центральности узлов сети цитирования как сумму рангов узлов сети по каждой мере центральности с учетом их релевантности.

Этап 4. Ранжирование публикаций по важности согласно профилю исследования

Ранжирование публикаций происходит в *четвертом блоке* модуля (2), где реализован метод ранжирования публикаций по важности, который сводится к вычислению суммы рангов узлов сети по каждой мере центральности с учетом их релевантности. Данный метод можно также назвать методом комплексной оценки центральности узлов сети – определяется значение комплексного показателя C^h (мера центральности по важности соответствующего профиля исследования). Далее осуществляется ранжирование публикаций в порядке убывания значения C^h .

Этап 5. Визуализация данных о цитировании

В СППР входит модуль (3) для просмотра и визуального анализа графов сетей цитирования на основе исследования и последующего ранжирования публикаций по важности. В качестве входной информации используются файл с библиографическими данными публикаций в виде таблицы MS Excel и файл с ориентированным графом в формате Rajek. Программный модуль выполняет следующие функции: выбор алгоритма укладки графа, обеспечивающего круговую, планарную, случайную, спектральную и специальные (для больших графов) визуализации; масштабирование вершины графа с учетом ранга важности узлов сети; отображение информации о публикации; создание HTML-файла визуального представления графа.

Этап 6. Оценка эффективности генератора списка источников информации в сетях цитирования

С целью подтверждения эффективности предложенного метода ранжирования публикаций по важности были проведены экспериментальные исследования: собраны и обработаны данные из различных баз цитирования RePEc, eLIBRARY.RU, Scopus, OpenAlex; результаты, полученные в ходе экспериментов, сравнены с результатами, полученными известными методами.

С помощью методов ранговой корреляции Спирмена и Стьюдента установлена тесная положительная линейная зависимость между значениями показателя, определенными с помощью метода ранжирования публикаций по важ-

ности, разработанного автором, и значениями мирового показателя цитируемости, взвешенного по предметной области – Field-Weighted Citation Impact (FWCI). Установлено также, что теснота связи между рангами, полученными по разработанному методу, и рангами, полученными по FWCI, зависит от профиля исследования (набора критериев – мер центральности узлов сети). Значение коэффициента ранговой корреляции в сравнении с профилем исследования «Входящий/Исходящий», рассмотренным в статье [26], свидетельствует о сильной прямой связи между двумя ранжированиями при критическом значении 0,01. При этом значение коэффициента ранговой корреляции в сравнении с профилем исследования «Входящий», а также с мерой центральности по полустепени захода (количеству цитирований научных публикаций), свидетельствует о слабой прямой связи при критическом значении 0,05. В итоге сделан вывод о целесообразности использования предложенной нами формулы определения важности узлов сети цитирования с помощью комбинирования сетевых метрик для улучшения и расширения показателей оценки научных публикаций.

Произведено сравнение результатов ранжирования научных статей разработанным методом с результатами, полученными методом анализа иерархий (МАИ). На основе полученных значений глобальных приоритетов МАИ проранжированы статьи-кандидаты. По результатам ранжирования статей двумя сравниваемыми методами установлено, что значения рангов R^c и $R^{МАИ}$ равны, где R^c – ранг статьи, полученный разработанным методом, $R^{МАИ}$ – ранг статьи по методу анализа иерархий. Следовательно, ранжирование разработанным методом произведено верно. Проведенные исследования позволили установить, что использование метода сетевого анализа научных публикаций для ранжирования статей является эффективным. Полученные результаты позволяют оценить согласованность рангов статей сравниваемых методов [21].

Стоит отметить, что метод ранжирования публикаций по важности позволяет работать с большим числом публикаций, что дает возможность улавливать более сложные закономерности и зависимости данных. Таким образом, очевидно преимущество применения этого метода на больших выборках в отличие от МАИ. Метод анализа иерархий имеет ограничения в применении на больших

выборках, так как для учёта всех критериев решения необходимо получать очень большие объёмы точной и достоверной информации [22].

ЗАКЛЮЧЕНИЕ

В результате проведенных исследований спроектирована система поддержки принятия решений при выборе источников информации в интернете. Разработан программный комплекс, позволяющий проводить сбор данных о цитировании, предобработку, визуальный анализ графа сети цитирования на основе исследования и последующего ранжирования публикаций по важности. В целом программный комплекс способствует эффективному анализу научных публикаций в автоматизированном режиме и сокращает время на обработку данных.

В основу работы программного комплекса заложен разработанный метод ранжирования научных публикаций по важности, который позволяет работать с большим числом публикаций, что дает возможность улавливать более сложные закономерности и зависимости данных. Таким образом, очевидно преимущество применения этого метода на больших выборках.

Полученные результаты экспериментальных исследований показали, что применение разработанных алгоритмов и метода для решения задач ранжирования публикаций на основе анализа сетей цитирования уменьшают время, затраченное на обработку информации без использования данной технологии, улучшают эффективность поиска источников информации и позволяют оптимизировать процесс выбора источников информации по предметной области.

Благодарность

Автор выражает благодарность Людмиле Альбертовне Денисовой, д.т.н., профессору кафедры «Автоматизированные системы обработки информации и управления» Омского государственного технического университета.

СПИСОК ЛИТЕРАТУРЫ

1. *Sylvester J. J.* On an Application of the New Atomic Theory to the Graphical Representation of the Invariants and Covariants of Binary Quantics, with Three Appendices // *American Journal of Mathematics*. 1878. Vol. 1. no. 1. P. 64–104.

<https://doi.org/10.2307/2369436>

2. *König D.* Theory of finite and infinite graphs. Boston: Birkhäuser, 1990. 426 p. ISBN 978-1-4684-8973-6.

3. *Egghe L. Rousseau R.* Introduction to Informetrics: quantitative methods in library, documentation and information science. Amsterdam: Elsevier Science Publishers, 1990. 450 p. ISBN 0-444-88493-9.

4. *Бредихин С.В., Ляпунов В.М., Щербакова Н.Г.* Библиометрические сети научных статей и журналов. Ин-т вычислительной математики и математической геофизики СО РАН. Новосибирск: Изд-во ИВМиМГ СО РАН, 2021. 334 с. ISBN 978-5-901548-44-8.

5. *Freeman L.* A set measures of centrality based on betweenness // *Sociometry*. 1977. Vol. 40, No. 1. P. 35–41. <https://doi.org/10.2307/3033543>

6. *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.) / Ин-т приклад. мат. им. М.В. Келдыша РАН. М.: Изд-во ИПМ им. М.В. Келдыша РАН, 2021. С. 265–278.

7. *Бредихин С.В., Ляпунов В.М., Щербакова Н.Г., Юргенсон А. Н.* Параметры «Центральности» узлов сети цитирования научных статей // Проблемы информатики. 2016. № 1 (30). С. 39–57.

8. *Бредихин С.В., Ляпунов В.М., Щербакова Н.Г.* Мера важности научной периодики – «центральность по посредничеству» // Проблемы информатики. 2014. № 3. С. 53–63.

9. *Бредихин С.В., Ляпунов В.М., Щербакова Н.Г.* Ранжирование коллекции периодических изданий базы данных RePEc на основе метрик Eigenfactors // Проблемы информатики. 2014. № 1 (22). С. 36–42.

10. *Бредихин С.В., Ляпунов В.М., Щербакова Н.Г.* Мера «центральности» для ранжирования научных статей // Проблемы информатики. 2015. № 1. С. 55–64.

11. *Бредихин С.В., Ляпунов В.М., Щербакова Н.Г.* Параметры пар узлов сети цитирования научных статей // Проблемы информатики. 2016. № 2. С. 30–49.

12. *Ding Y., Yan E., Frazho A., Caverlee J.* PageRank for ranking authors in co-citation networks // *Journal of the American Society for Information Science and Technology*. 2009. Vol. 60, No. 11. P. 2229–2243. <https://doi.org/10.1002/asi.v60:11>

13. *Ding Y.* Applying weighted PageRank to author citation networks // Journal of the American Society for Information Science and Technology. 2011. Vol. 62, No. 2. P. 236–245. <https://doi.org/10.1002/asi.21452>

14. *Maslov S., Redner S.* Promise and pitfalls of extending Google's PageRank algorithm to citation networks // Journal of Neuroscience. 2008. Vol. 28. No. 44. P. 11103–11105.

15. *Jezek K., Fiala D., Steinberger J.* Exploration and Evaluation of Citation Networks // Open Scholarship: Authority, Community, and Sustainability in the Age of Web 2.0: Proceedings of the 12th International Conference on Electronic Publishing held (Toronto, Canada, 25–27 June 2008) / University of West Bohemia. Pilsen: University of West Bohemia, 2008. P. 351–362.

16. *Nykl M., Ježek K., Fiala D., Dostal M.* PageRank variants in the evaluation of citation networks // Journal of Informetrics. 2014. Vol. 8, No. 3. P. 683–692.

17. *Mejia C., Wu V., Zhang Y., Kajikawa Y.* Exploring topics in bibliometric research through citation networks and semantic analysis // Frontiers in Research Metrics and Analytics. 2021. Vol. 6. P. 742311.

18. *Ольгина И.Г., Бадрызлов С.В.* Анализ сетей цитирования научных публикаций // Системы управления, информационные технологии и математическое моделирование: материалы V Всерос. науч.-практ. конф. с междунар. участием (Омск, 25–26 апреля 2023 г.). Омск: ОмГТУ, 2023. С. 91–99.

19. Свидетельство о государственной регистрации программы для ЭВМ № 2020615709 Российская Федерация. Программный комплекс LinkAnalyzer 1.0 для сбора и анализа информации о цитировании научных публикаций: № 2020614572: заявл. 21.05.2020: опубл. (зарег.) 29.05.2020 / И.Г. Ольгина, А.Н. Абдрахманов; заявитель Ом. гос. техн. ун-т. 1 с.

20. Свидетельство о государственной регистрации программы для ЭВМ № 2021661693 Российская Федерация. Генератор списка источников информации в сетях цитирования: № 2021660969: заявл. 14.07.2021: опубл. (зарег.) 14.07.2021 / И.Г. Ольгина, Д.С. Осипов; заявитель Ом. гос. техн. ун-т. 1 с.

21. *Ольгина И.Г.* Анализ результатов ранжирования научных публикаций с использованием метода ранговой корреляции // Известия Тульского государственного университета. Технические науки. 2024. № 3. С. 74–78.

<https://doi.org/10.24412/2071-6168-2024-3-74-75>

22. *Ольгина И.Г., Андреева Е.Г., Янишевская А.Г., Федоров В.К.* Сравнительный анализ метода сетевого анализа научных публикаций и метода анализа иерархий при выборе источников информации // Известия Тульского государственного университета. Технические науки. 2024. № 9. С. 484–488.

<https://doi.org/10.24412/2071-6168-2024-9-484-485>

23. Свидетельство о государственной регистрации программы для ЭВМ № 2023666387 Российская Федерация. Визуализатор сетей цитирования: № 2023665592; заявл. 27.07.2023; опубл. 31.07.2023/ И.Г. Ольгина, С.В. Бадрызов; заявитель и правообладатель Ом. гос. техн. ун-т. 1 с.

24. *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.

25. RePEC. General principles. URL: <http://repec.org/> (дата обращения: 26.11.2024).

26. *Ольгина И.Г.* Программные и алгоритмические средства сетевого анализа для определения важности научных публикаций // Электронные библиотеки. 2023. № 5. С. 646–672. <https://doi.org/10.26907/1562-5419-2023-26-5-646-672>

SUPPORT SYSTEM FOR THE SELECTION OF INFORMATION SOURCES IN CITATION NETWORKS

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

Omsk State Technical University

inna_olgina@mail.ru

Abstract

With the advent of network science, it has become possible to explore complex network systems, including social and information networks, by presenting them as graph models. The exponential growth of the total volume of scientific publications determines the relevance of the tasks of analyzing their interrelations. In network

science, models and methods related to the field of so-called citation networks are being developed to solve these problems. However, network metrics are not used when analyzing publications in citation databases. The paper considers the issues of creating a decision support system for the selection of information sources based on data on the citation of scientific publications. A software package has been developed for making decisions on determining an important publication in a certain thematic area. The software package is based on a method of ranking publications by importance based on the analysis of citation networks, which allows you to identify publications that do not clearly stand out when ranking based on known bibliometric indicators or known measures of centrality of nodes in their pure form. A study and comparative analysis of software for visualization and research of all types of graphs and social networks has been conducted. Studies have been carried out confirming the effectiveness of the proposed decision support system in the selection of information sources.

Keywords: *citation network, publication, scientometry, decision support system, software architecture, network analysis, graph.*

REFERENCES

1. Sylvester J.J. On an Application of the New Atomic Theory to the Graphical Representation of the Invariants and Covariants of Binary Quantics, with Three Appendices // American Journal of Mathematics. 1878. Vol. 1, No. 1. P. 64–104. <https://doi.org/10.2307/2369436>
2. König D. Theory of finite and infinite graphs. Boston: Birkhäuser, 1990. 426 p. ISBN 978-1-4684-8973-6.
3. Egghe L. Rousseau R. Introduction to Informetrics: quantitative methods in library, documentation and information science. Amsterdam: Elsevier Science Publishers, 1990. 450 p. ISBN 0-444-88493-9.
4. Bredihin S.V., Ljapunov V.M., Shherbakova N.G. Bibliometric networks of scientific articles and journals. Institute of Computational Mathematics and Mathematical Geophysics, Siberian Branch RAS: ICM&MG SB RAS, 2021. 334 p. ISBN 978-5-901548-44-8. (In Russ.).

5. *Freeman L.* A set measures of centrality based on betweenness // *Sociometry*. 1977. Vol. 40, No. 1. P. 35–41. <https://doi.org/10.2307/3033543>
 6. *Pechnikov A.A., Chebukov D.E.* 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.).
 7. *Bredihin S.V.* Centrality parameters of nodes of the scientific articles citation network // *Problemy Informatiki – Problems of Informstics*. 2016. No. 1. P. 39–57 (In Russ.).
 8. *Bredihin S.V., Ljapunov V.M., Shherbakova N.G.* Measurement of the scientific periodicals importance: cooperation centrality // *Problemy Informatiki – Problems of Informstics*. 2014. No. 3. P. 53–64 (In Russ.).
 9. *Bredihin S.V., Ljapunov V.M., Shherbakova N.G.* Ranking of the collection of periodicals in the RePEc database based on Eigenfactor metrics // *Problemy Informatiki – Problems of Informstics*. 2014. No. 1 (22). P. 36–42 (In Russ.).
 10. *Bredihin S.V., Ljapunov V.M., Shherbakova N.G.* A measure of "centrality" for ranking scientific articles // *Problemy Informatiki – Problems of Informstics*. 2015. No. 1. P. 55–64 (In Russ.).
 11. *Bredihin S.V., Ljapunov V.M., Shherbakova N.G.* Parameters of pairs of nodes of the network for quoting scientific articles // *Problemy Informatiki – Problems of Informstics*. 2016. No. 2. P. 30–49 (In Russ.).
 12. *Ding Y., Yan E., Frazho A., Caverlee J.* PageRank for ranking authors in co-citation networks // *Journal of the American Society for Information Science and Technology*. 2009. Vol. 60, No. 11. P. 2229–2243. <https://doi.org/10.1002/asi.v60:11>
 13. *Ding Y.* Applying weighted PageRank to author citation networks // *Journal of the American Society for Information Science and Technology*. 2011. Vol. 62, No. 2. P. 236–245. <https://doi.org/10.1002/asi.21452>
 14. *Maslov S., Redner S.* Promise and pitfalls of extending Google's PageRank algorithm to citation networks // *Journal of Neuroscience*. 2008. Vol. 28. No. 44. P. 11103–11105.
 15. *Jezek K., Fiala D., Steinberger J.* Exploration and Evaluation of Citation Networks // *Open Scholarship: Authority, Community, and Sustainability in the Age of*
-

Web 2.0: Proceedings of the 12th International Conference on Electronic Publishing held (Toronto, Canada, 25–27 June 2008) / University of West Bohemia. Pilsen: University of West Bohemia, 2008. P. 351–362.

16. *Nykl M., Ježek K., Fiala D., Dostal M.* PageRank variants in the evaluation of citation networks // *Journal of Informetrics*. 2014. Vol. 8, No. 3. P. 683–692.

17. *Mejia C., Wu V., Zhang Y., Kajikawa Y.* Exploring topics in bibliometric research through citation networks and semantic analysis // *Frontiers in Research Metrics and Analytics*. 2021. Vol. 6. P. 742311.

18. *Olgina I.G., Badryzlov S.V.* Analysis of citation networks of scientific publications]. *Sistemy upravlenija, 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. 2023. P. 91–99 (In Russ.).

19. *Olgina I.G., Abdrahmanov A.N.* Certificate of state registration of the computer program No. 2020615709 Russian Federation. Link Analyzer 1.0 software package for collecting and analyzing information about citation of scientific publications: No. 2020614572: application 21.05.2020: publ. 29.05.2020; the applicant is the Omsk State Technical University.

20. *Olgina I.G., Osipov D.S.* Certificate of state registration of the computer program No. 2021661693 Russian Federation. Generator of a list of information sources in citation networks: No. 2021660969: application 14.07.2021: publ. 14.07.2021; the applicant is the Omsk State Technical University.

21. *Olgina I.G.* Analysis of the results of ranking scientific publications using the rank correlation method // *News of the Tula state university. Technical sciences*. 2024. No. 3. P. 74–78 (In Russ.). <https://doi.org/10.24412/2071-6168-2024-3-74-75>

22. *Olgina I.G., Andreeva E.G., Yanishevskaya A.G., Fedorov V.K.* Comparative analysis of the method of network analysis of scientific publications and the method of hierarchy analysis in the selection of information sources // *News of the Tula State University. Technical sciences*. 2024. No. 9. P. 484–488 (In Russ.). <https://doi.org/10.24412/2071-6168-2024-9-484-485>

23. *Olgina I.G., Badryzlov S.V.* Certificate of state registration of the computer program No. 2023666387 Russian Federation. Citation Network Visualizer:

No. 2023665592; application 27.07.2023; publ. 31.07.2023; the applicant is the Omsk State Technical University.

24. *Olgina I.G., Pronin I.V., Abdrahmanov A.N.* Graph models construction of the citation network of scientific publications // *Sistemy upravlenija, 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. V. 1. P. 118–125 (In Russ.).

25. RePEC. General principles. URL: <http://repec.org/> (date accessed: 26.11.2024).

26. *Olgina I.G.* Software and algorithmic tools of network analysis for determining the scientific publications importance // *Russian Digital Libraries Journal. 2023. No. 5. P. 646–672.* <https://doi.org/10.26907/1562-5419-2023-26-5-646-672>.

СВЕДЕНИЯ ОБ АВТОРЕ



ОЛЬГИНА Инна Геннадьевна – старший преподаватель кафедры «Математические методы и информационные технологии в экономике», ОмГТУ. Область научных интересов: теория графов, применение методов системного анализа для исследования сетей цитирования, цифровые технологии в работе библиотек. Число научных публикаций – 21.

Inna Gennadevna OLGINA – 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, digital technologies in the work of libraries. The number of publications – 21.

email: inna_olgina@mail.ru

ORCID: 0000-0002-9932-4552

Материал поступил в редакцию 28 ноября 2024 года