

ВЫЯВЛЕНИЕ МЕХАНИЗМОВ НЕДОБРОСОВЕСТНОГО ПОВЫШЕНИЯ НАУКОМЕТРИЧЕСКИХ ПОКАЗАТЕЛЕЙ

Емельянов Д. С.¹, студент, ✉ dima31120251@gmail.com
Черепанов Р. П.^{1,2}, студент, romacherepanov2002@gmail.com

¹ Санкт-Петербургский государственный электротехнический университет «ЛЭТИ»
им. В. И. Ульянова (Ленина), ул. Профессора Попова, 5, корп. 3, 197022, Санкт-Петербург, Россия

² ООО 1С-Софт, Дмитровское ш., д. 9, 127434, Москва, Россия

Аннотация

Цель исследования состоит в выявлении коллективов авторов, недобросовестно повышающих свои наукометрические показатели путем частого взаимного цитирования. Если составить граф цитирований, где вершинами будут выступать авторы, а ребрами — отношения «автор процитировал автора», то такими подозрительными коллективами могут быть клики, на которые есть мало ссылок от других авторов, а внутренних ссылок, наоборот, много.

Были получены метаданные о научных публикациях из сервиса Crossref, проведена их фильтрация и приведение к удобному для обработки виду с последующей загрузкой в локальную базу данных MySQL. После этапа переноса данных были выделены неориентированный граф цитирований и его компоненты связности, найдены максимальные клики и проведён анализ их статистических характеристик. С учетом этих характеристик выделены наиболее подозрительные коллективы. Данный метод может быть использован для поиска групп авторов, цитирующих друг друга по договоренности.

Ключевые слова: *Crossref, наукометрия, социальный граф, клика, алгоритм Брона-Кербоша, прикладные задачи на графах, базы данных.*

Цитирование: Емельянов Д. С., Черепанов Р. П. Выявление механизмов недобросовестного повышения наукометрических показателей // Компьютерные инструменты в образовании. 2023. № 1. С. 18–29. doi:10.32603/2071-2340-2023-1-18-29

1. ВВЕДЕНИЕ

Использование системы наукометрии, включая разнообразные индексы цитируемости, и поощрение наукометрических показателей неизбежно привело к появлению авторов и коллективов, эксплуатирующих эту систему. Такими коллективами могут быть несколько авторов, ссылающихся друг на друга по договоренности. Выявление таких групп авторов позволит бороться с недобросовестным повышением индексов цитируемости, что является актуальной проблемой.

Данная тема исследуется с помощью разных подходов. В [1] для оценки научного значения работ введен показатель «Conflict Of Interests». Данный показатель учитывает аффилированность авторов и совместные публикации, а также время, в течение которого

эти авторы совместно публиковались. С учетом этого показателя вычисляется суммарный «вес» работы, который зависит от количества ссылок на нее, где ссылки от авторов, тем или иным образом связанных с авторами публикации, имеют меньшее значение.

В работе [2] посредством анализа графа цитирований между журналами, тематика которых — компьютерные науки, выделены наборы журналов разных типов: пары, тройки, цепи, группы и картели цитирований. Сначала исключаются журналы, менее других участвующих в цитировании, затем полученный граф исследуется. С учетом самоцитирований и времени цитирования оценка научного значения журналов пересчитывается.

В статье [3] задача поиска недобросовестных цитирований сведена к задаче классификации журналов с учетом количества самоцитирований журнала, общего количества ссылок на него и количества опубликованных работ.

В публикации [4] представлен алгоритм «CIDRE», направленный на выявление наборов научных журналов, наиболее часто ссылающихся друг на друга. Для этого учитываются ссылки из одного журнала на статьи другого, написанные в течение последних двух лет. Полученные данные сравниваются с наборами цитирований, случайно сгенерированными по определенным правилам. Те наборы журналов, где количество цитирований превосходит случайно сгенерированное, обрабатываются алгоритмом, и далее из них выделяются наиболее подозрительные группы журналов.

В работе 2022 года [5] представлен инструмент «GLAD» — модель машинного обучения, которая изучает тексты работ и определяет, уместна ли та или иная ссылка с точки зрения контекста.

Наша работа касается именно поиска потенциально недобросовестных коллективов авторов, а не оценки влияния отдельной работы или журнала. Представив данные об авторах в виде графа цитирований, можно найти клики авторов и узнать, работы каких коллективов цитируются чаще или реже. Авторы, привлекающие мало чужого внимания, но при этом очень часто цитирующие других участников клики, вероятно, недобросовестно повышают друг другу индексы цитирования. Поиск таких коллективов является целью работы.

2. ПОСТРОЕНИЕ И РЕАЛИЗАЦИЯ МАТЕМАТИЧЕСКОЙ МОДЕЛИ

Чтобы построить граф цитирований, необходим массив библиографических данных о различных работах, их авторах и ссылках между работами. Такие данные предоставляются различными специализированными сервисами.

В исследовании используется дамп базы данных сервиса Crossref 2020 года [6]. Crossref — это официальное агентство регистрации Цифровых Идентификаторов Объекта международного DOI фонда, которое предоставляет доступ к необходимым в нашей работе метаданным на свободной основе.

Дамп базы данных представляет собой архив из файлов с расширением «.json», в которых содержатся записи с различной информацией о каждой работе. Суммарно архив весит 60 ГБ и содержит примерно 35 000 файлов по 5000 записей о работах.

Работа с архивом делится на следующие этапы:

1. Извлечение информации. Для обработки используются только корректные и релевантные данные.
2. Построение графа цитирований. По данным о публикациях составляется граф цитирований «автор ссылается на автора».
3. Поиск клик с помощью алгоритма Брона-Кербоша [7].

4. Нахождение потенциально подозрительных коллективов с учетом количества ссылок между их работами и ссылок из остальных работ.

В исследовании используются работы, опубликованные не ранее 2000 года и не позднее мая 2020 года. Код проекта написан на языке программирования Python 3.10. Локальная база данных, в которой сохраняются результаты выполнения каждого из этапов, реализована с помощью СУБД MySQL.

2.1. Выделение релевантной информации

Запись в базе содержит большое количество данных о статье, но для построения графа интересна лишь небольшая часть из них: уникальный идентификатор работы DOI, список авторов, список ссылок на другие работы, область науки, которой посвящена работа, год публикации, количество цитирований данной работы из других работ.

Для сокращения объема обрабатываемой информации в рамках исследования выбирается одна из областей науки, статьи по которой есть в дампе, или несколько смежных областей. В частности, в данной работе рассматриваются только те статьи, чьи записи в поле «subject» содержат значение «Electrical and Electronic Engineering».

Чтобы выделить из дампа необходимую информацию, записи о работах последовательно проверяются на соответствие выбранной области науки и году публикации. Также выполняется проверка на корректность, включающая в себя проверку наличия необходимых полей:

- **DOI** — цифровой идентификатор работы,
- **author** — список авторов,
- **reference** — список ссылок,
- **published-print** — год публикации,
- **subject** — список областей науки, к которым относится работа,
- **is-referenced-by-count** — количество цитирований данной работы из других работ.

Проверенные записи с этими полями сохраняются в новые файлы, остальные поля не используются.

2.2. Проблема идентификации авторов

Существует проблема с идентификацией авторов в научной среде. В работах авторы представлены без уникальных идентификаторов, поименно. В 2011 году появился идентификатор авторов — ORCID [8], однако подавляющее большинство работ не содержит ни одного автора с ORCID. Мы предполагаем, что авторы с полностью совпадающими именами — один и тот же человек. Кроме того, если два автора с одинаковой фамилией и одинаковыми инициалами ссылаются друг на друга, но у одного из них имя записано сокращенно, считается, что это один человек, если такая связь одна. Например, если «Н. Potter» ссылается на «Harry Potter» и больше нет никаких ссылок на авторов с именем вида «Н* Potter», тогда предполагается, что эти два автора — один и тот же человек. Поэтому в работе реализован отдельный этап слияния авторов. Пример такой ситуации изображен на рисунке 1, где вершины соответствуют авторам, а взвешенные ребра — количеству ссылок от одного автора на другого.

Автора «Н. Potter» можно объединить с автором «Harry Potter», так как ссылка определена однозначно: «Н. Potter» не ссылается на «Henry Potter» и наоборот.

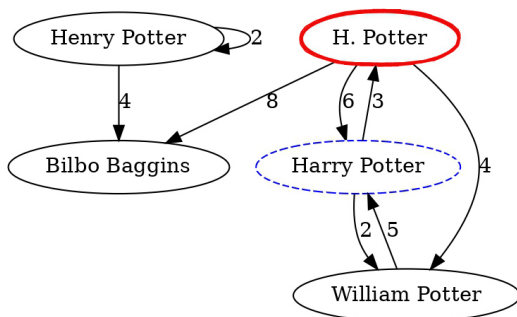


Рис. 1. Однозначная ссылка автора с неполным именем (жирная линия) на автора с полным (пунктирная линия)

2.3. Построение графа цитирований и нахождение клик

Записи в созданных на предыдущем этапе файлах обрабатываются дважды.

На первой итерации в локальную БД записываются следующие данные: имена авторов, DOI, год написания работы, отношения «автор написал работу». На второй итерации в БД сохраняются отношения цитирования: «автор ссылается на автора» и «работа ссылается на работу». Дальнейшая работа производится над таблицей цитирований «автор ссылается на автора», которая представляет собой граф с авторами-вершинами и взвешенными ребрами — количеством ссылок одного автора на другого, который для краткости называется ориентированным графом цитирований. Пример графа представлен на рисунке 2, где числа в вершинах — это ID авторов в БД. Ребра приведены без весов для удобства восприятия.

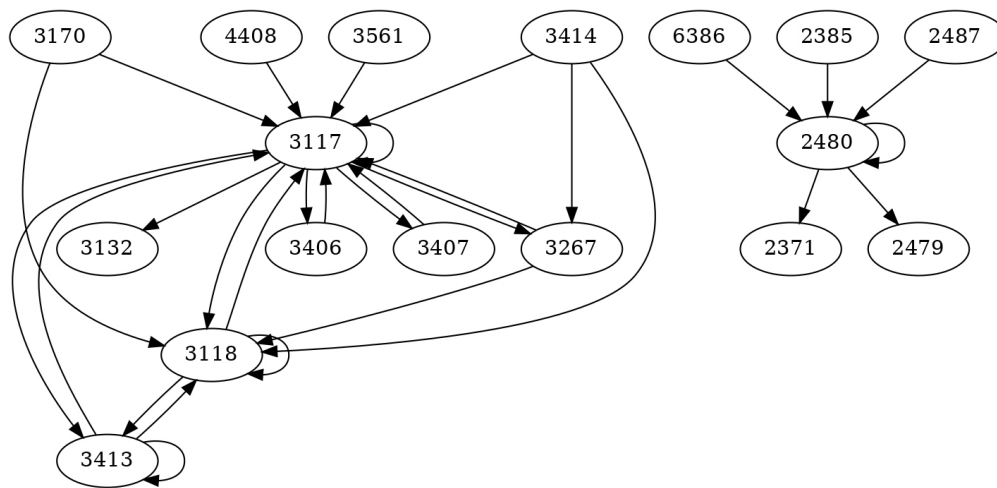


Рис. 2. Ориентированный граф цитирований

После записи в базе данных происходит слияние авторов по принципу, описанному выше: считается, что если есть два автора, имя одного из которых записано инициалами и между ними есть ссылка, то это один человек, поэтому записи в БД, касающиеся этих авторов, обновляются.

Далее выделяется неориентированный граф цитирований. В нем сохраняются только двунаправленные ребра, то есть только взаимные цитирования. Кроме того, можно огра-

ничить минимальные веса ребер из начального графа, которые могут образовать ребра в новом графе, то есть наложить ограничение, которое будет учитываться при поиске клик, на минимальное количество взаимных цитирований. В рассмотренной области науки это ограничение не применялось. Вид неориентированного графа, выделенного из графа на рисунке 2, представлен на рисунке 3.

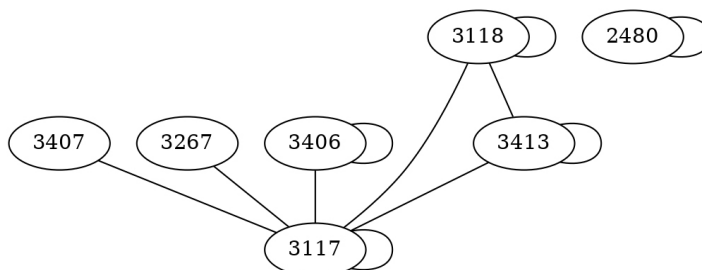


Рис. 3. Неориентированный граф цитирований

Эти ограничения позволяют значительно ускорить процесс обработки, уменьшив количество ребер, однако самые недобросовестные коллективы авторов по-прежнему будут выделены, так как количество цитирований между собой у них, предположительно, достаточно велико.

На полученном графе с помощью поиска в ширину выделяются компоненты связности — это позволяет разделить задачу обработки на меньшие части. На рисунке 3 есть две компоненты связности.

После выделения компонент связности в каждой компоненте производится поиск максимальных клик с помощью нашей реализации алгоритма Брона-Кербоша [7]. Данный алгоритм является детерминированным, его асимптотическая сложность по времени составляет $O(3^{\frac{n}{3}})$, где n — количество вершин графа. Каждая максимальная по включению авторов клика сохраняется в БД. На рисунке 4 изображена одна из двух компонент графа из рисунка 3, в данной компоненте находятся четыре максимальные клики:

- 3407, 3117;
- 3267, 3117;
- 3406, 3117;
- 3413, 3118, 3117.

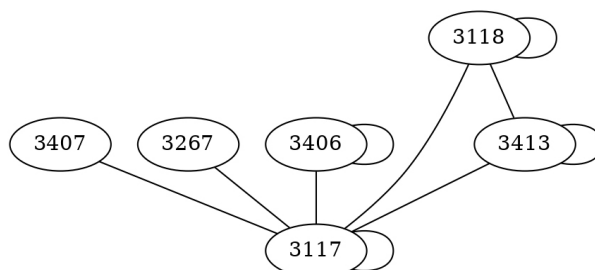


Рис. 4. Одна из компонент связности неориентированного графа цитирований

2.4. Выделение подозрительных коллективов

Для оценки степени подозрительности клик используются коэффициенты внутреннего и внешнего цитирования. Для их определения введем следующие понятия:

- **работа клики** — работа, хотя бы один из авторов которой состоит в данной клике;
- **внутренняя ссылка клики** — ссылка от одного автора клики на другого автора клики (самоцитирования не учитываются).

Коэффициент внутреннего цитирования (*internal_citing*) должен показывать, как часто авторы клики ссылаются друг на друга. В рамках исследования в качестве такого коэффициента используется отношение количества внутренних ссылок (*internal_refs*) к квадрату размера клики (*clique_size*):

$$\text{internal_citing} = \frac{\text{internal_refs}}{\text{clique_size}^2} \quad (1)$$

Например, пусть имеется клика из 3 человек, всего работ у клики — 2, существует ссылка из работы 1 на работу 2, причем в обеих работах авторами являются члены клики. Тогда внутренних ссылок у этой клики будет 6: первый автор сослался на второго и третьего авторов, второй — на первого и третьего, третий — на первого и второго.

Коэффициент внешнего цитирования (*external_citing*) должен отражать интерес других ученых к работам клики и является показателем, характеризующим клику с положительной стороны. В качестве коэффициента внешнего цитирования используется отношение количества всех ссылок (*refs*) к количеству работ клики (*clique_works*). Количество всех ссылок подсчитывается суммированием значения «is-referenced-by-count» всех работ клики.

$$\text{external_citing} = \frac{\text{refs}}{\text{clique_works}} = \frac{\sum \text{is-referenced-by-count}}{\text{clique_works}} \quad (2)$$

При подсчете коэффициентов должны учитываться и односторонние цитирования, то есть подсчет коэффициентов производится на ориентированном графе цитирований, полученном после слияния авторов, но до выделения неориентированного графа.

Коэффициенты одной клики дают мало информации сами по себе, так как в разных областях науки среднее количество ссылок в работе и количество работ у автора различаются. Но, вычислив коэффициенты цитирования многих клик, можно сравнить между собой, например, отношение коэффициента внешнего цитирования к внутреннему и получить относительный показатель. Чем меньше этот показатель, тем более подозрительной является клика. Отсортировав клики по этому отношению, можно найти потенциально недобросовестные коллективы авторов, а затем по БД восстановить список их работ.

Повторное исследование можно провести на том же графе, но с дополнительным условием: в клику включаются только соавторы, то есть коллективы, у каждой пары авторов из которых есть хотя бы одна совместная работа. Количество клик и их коэффициенты изменятся, однако самые подозрительные коллективы можно выделить таким же образом.

3. РЕЗУЛЬТАТЫ

Был проведен анализ работ из области «Electrical and Electronic Engineering». Сначала были выделены клики, у которых коэффициент внутреннего цитирования больше 1 и количество работ больше 10. Затем эти клики были упорядочены по неубыванию отношения коэффициента внешнего цитирования к внутреннему. Таким образом, наиболее подозрительные клики расположились в начале списка. В таблице 1 представлена информация о первых пяти коллективах из этого списка, причем под «внутренней ссылкой клики» подразумевается понятие, введенное в пункте 2.4.

Таблица 1. Информация о пяти самых подозрительных кликах

№	Размер	Количество работ	Ссылки на работы	Внутренние ссылки клики
1	4	13	19	124
2	6	11	6	50
3	3	15	70	100
4	3	15	70	100
5	3	15	70	100

Для проверки полученных результатов нужна более актуальная информация о работах, так как они могли получить много новых ссылок за период с мая 2020 года до момента проведения анализа данных клик, то есть ноября 2022 года. Для этого используется сервис [9].

Клика №1 была проверена вручную. Путем просмотра информации о записанных в БД работах этой клики с помощью указанного сервиса были получены актуальные данные о ссылках. Для каждой работы клики было получено общее количество ссылок на нее и количество ссылок от работ этой же клики. График полученных данных изображен на рисунке 5.

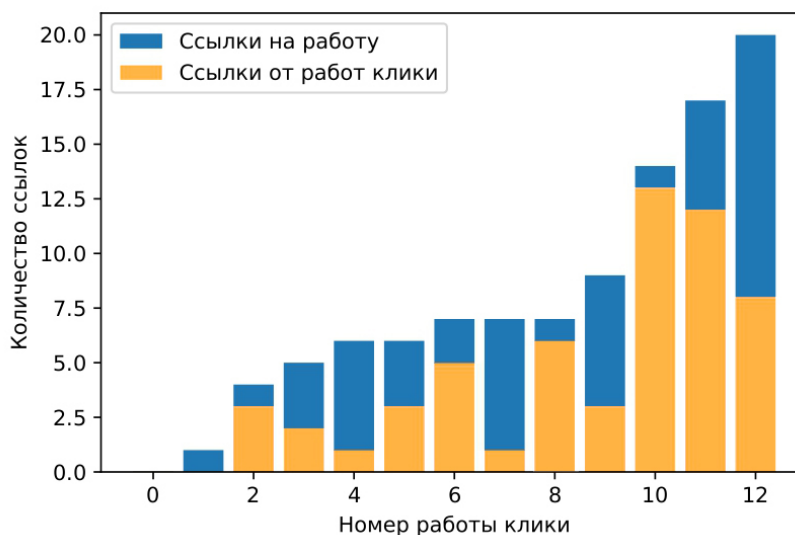


Рис. 5. Количество ссылок на работы клики №1

Из рисунка 5 видно, что многие из работ коллектива достаточно часто цитируются теми же авторами. Как минимум в 6 из 12 рассмотренных работ более половины цитирований приходится на работы клики.

Авторы были распознаны алгоритмом как подозрительные, так как в базе данных за исследуемый период у них было мало ссылок, причем большинство из них были от участников клики (в подавляющем большинстве ссылок фигурирует один и тот же автор). С этой точки зрения коллектив действительно подозрительный: у клики выходит много работ, в которых большинство ссылок — самоцитирования.

Таким же образом была проанализирована клика № 2, но она оказалась не такой подозрительной, как первая. Причина, по которой этот коллектив расположился в начале списка, заключается в том, что работы клики были опубликованы не так давно, из-за чего в БД количество внешних ссылок на них достаточно мало. На рисунке 6 изображен график с актуальными данными о ссылках (в начале идут работы с нулевым цитированием).

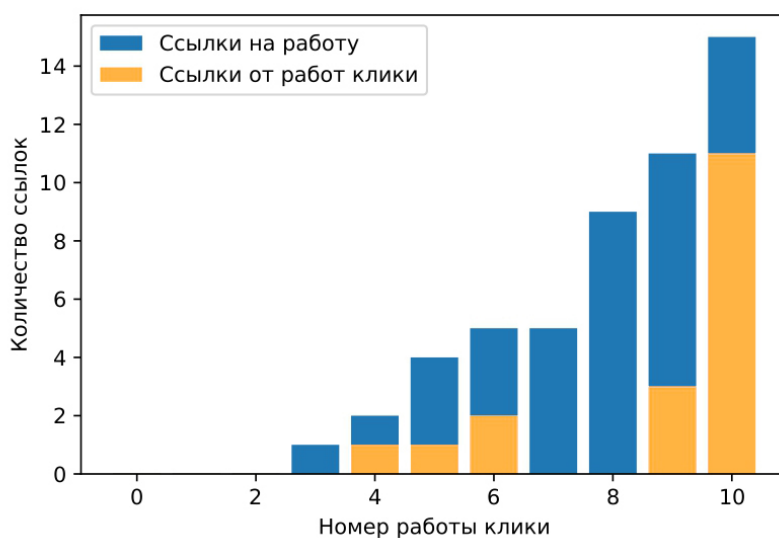


Рис. 6. График количества ссылок на работы клики № 2

Клики № 3–36 имеют размер 3, причем в каждой из них числится один и тот же автор. График с актуальными данными для клики № 3 представлен на рисунке 7, остальные коллективы из этого набора не представляют особого интереса, так как их коэффициенты цитирования во многом определены наличием этого общего автора.

Отдельно были рассмотрены группы авторов с нулевым внешним цитированием, однако для них коэффициенты внутреннего цитирования оказались меньше 1, что говорит о достаточно слабой связи авторов внутри этих клик.

Итого, описанный метод позволяет выделять подозрительные коллективы авторов, но нельзя точно сказать, повышают ли они свои индексы цитируемости недобросовестно. Во-первых, для получения более точных данных нужно рассматривать несколько смежных областей науки, так как, например, работа из «Астрономии» может ссылаться на работу из «Общей физики». Во-вторых, есть вероятность полного совпадения имен у двух разных авторов, из-за чего они будут считаться одним человеком. Наконец, некоторые коллективы могут работать в узкой сфере, где выпускается мало работ, из-за чего цитирование от других авторов значительно уменьшается.

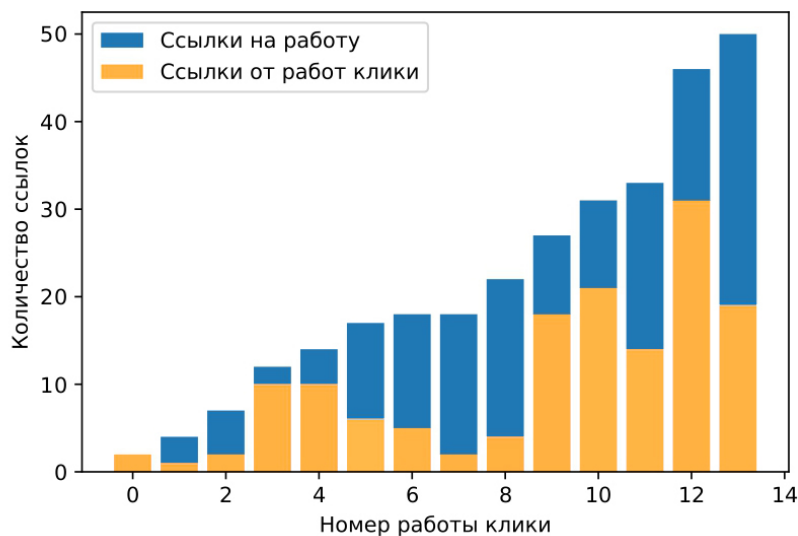


Рис. 7. График количества ссылок на работы клики № 3

Несмотря на эти недостатки, данный метод позволяет выделить наиболее подозрительные коллективы, которые в дальнейшем можно проверить более точными способами за меньшее количество времени, так как объем рассматриваемых данных существенно сократится.

4. ЗАКЛЮЧЕНИЕ

Представленный в данной работе метод позволяет выявить коллективы, недобросовестно повышающие свои индексы, путем представления массива библиографических данных в виде графа цитирований и его дальнейшего анализа. Авторы в графе представляют собой вершины, а ссылки между ними — ребра. Отбрасываются однонаправленные ребра, выделяются компоненты связности. В компонентах связности ищутся клики авторов, после чего происходит вычисление их коэффициентов цитирования. Далее клики сортируются по невозрастанию меры подозрительности, основанной на этих коэффициентах. Таким образом, клики в начале списка являются самыми подозрительными и могут быть изучены подробнее другими способами.

Для реализации данного метода была написана программа на Python 3.10, использующая для хранения данных СУБД MySQL 8.0. После этапа фильтрации количество работ составило 89533. Размер графа цитирований составил 223723 узлов-авторов и 637527 ребер-ссылок. Всего было найдено 13389 кликов, содержащих 34245 авторов. Три наиболее подозрительные из них были проверены вручную с учетом актуальной информации. Среди проверенных кликов первая и третья имеют подозрительно большое количество внутренних ссылок по отношению к их общему количеству.

Рассмотренный метод не позволяет сделать однозначный вывод о недобросовестном цитировании в коллективе, но может быть использован как инструмент для нахождения подозрительных коллективов.

Проект можно дорабатывать для получения более точных результатов и увеличения производительности. Код некоторых этапов может быть переписан на C++ для возможного ускорения работы. Для лучшей оценки подозрительности можно ввести показатель

самоцитирования автора, который будет учитываться при оценке клики, в которой он состоит. Для определения подозрительности коллектива могут использоваться другие показатели или другие их соотношения. В будущем могут быть обработаны другие области науки с большим количеством работ в базе данных Crossref, например, «Астрономия».

Код проекта и схема базы данных доступны в репозитории [10].

Список литературы

1. Bai X., Xia F., Lee I., Zhang J., Ning Z. Identifying Anomalous Citations for Objective Evaluation of Scholarly Article Impact // PLOS ONE. 2016. Vol. 11, № 9. P. e0162364. doi:10.1371/journal.pone.0162364.
2. Chakraborty, J., Pradhan D., Dutta H. S., Nandi S., Chakraborty T. On Good and Bad Intentions behind Anomalous Citation Patterns among Journals in Computer Sciences. arXiv, 2018. doi:10.48550/ARXIV.1807.10804.
3. Yu T., Yu G., Song Y., Wang M.-Y. Toward the more effective identification of journals with anomalous self-citation // Malaysian Journal of Library & Information Science. 2018. Vol. 23, № 2. P. 25–46. doi: 10.22452/mjlis.vol23no2.2.
4. Kojaku S., Livan G., Masuda N. Detecting anomalous citation groups in journal networks // Scientific Reports. 2021. Vol. 11. P. 14524. doi: 10.1038/s41598-021-93572-3.
5. Liu J., Xia F., Feng X., Ren, J., Liu H. . Deep Graph Learning for Anomalous Citation Detection // IEEE Transactions on Neural Networks and Learning Systems. 2022. Vol. 33, № 6. P. 2543–2557. doi: 10.1109/TNNLS.2022.3145092.
6. Kemp, J. Free public data file of 112+ million Crossref records. [Online]. URL: <https://www.crossref.org/blog/free-public-data-file-of-112-million-crossref-records/> (date: 16.03.2023).
7. Bron C., Kerbosch J. Algorithm 457: finding all cliques of an undirected graph // Communications of the ACM. 1973. Vol. 16, № 9. P. 575–577. doi:10.1145/362342.362367
8. ORCID Inc. About ORCID. 2023. [Online]. URL: <https://info.orcid.org/> (date: 16.03.2023).
9. Digital Science & Research Solutions Inc. The data in Dimensions — from idea to impact. 2023. [Online]. URL: <https://www.dimensions.ai/dimensions-data/> (date: 16.03.2023).
10. D. Emelyanov. Cliques In Science. 2022. [Online]. URL: https://github.com/Deimos-Apollon/Cliques_In_Science (date: 16.03.2023).

Поступила в редакцию 17.01.2023, окончательный вариант — 16.03.2023.

Емельянов Дмитрий Сергеевич, студент 3-го года обучения образовательной программы бакалавриата «Прикладная математика и информатика» СПбГЭТУ «ЛЭТИ»,
✉ dima31120251@gmail.com

Черепанов Роман Павлович, студент 3-го года обучения образовательной программы бакалавриата «Прикладная математика и информатика» СПбГЭТУ «ЛЭТИ»,
romacherepanov2002@gmail.com

Computer tools in education, 2023

№ 1: 18–29

<http://cte.eltech.ru>

[doi:10.32603/2071-2340-2023-1-18-29](https://doi.org/10.32603/2071-2340-2023-1-18-29)

Detection of Mechanisms for Unfair Increase of Scientometric Indexes

Emelyanov D. S.¹, Student, ✉ dima31120251@gmail.com
Cherepanov R. P.^{1,2}, Student, romacherepanov2002@gmail.com

¹Saint Petersburg Electrotechnical University,
5, building 3, st. Professora Popova, 197022, Saint Petersburg, Russia
²1C COMPANY, 9, Dmitrovskoe sh., 127434, Moscow, Russia

Abstract

The goal of the research is to detect groups of authors that increase their scientometric indexes through excessively frequent mutual citations. In a citation graph with vertices and edges representing authors and «author-cites-author» relations respectively, suspicious groups might be cliques that have few citations from other authors and, conversely, many internal citations. Metadata on scientific publications from the Crossref service was retrieved, filtered and reduced to an easy-to-process form, and then uploaded to the local MySQL database. After the data transfer phase the undirected citation graph and its connectivity components were extracted, the maximum cliques were found and their statistical characteristics were analyzed. The most suspicious collectives were detected using these characteristics. This method can be used to detect groups of authors who cite each other by agreement.

Keywords: *Crossref, scientometrics, social graph, clique, Bron-Kerbosch algorithm, applications of graph theory, databases.*

Citation: D. S. Emelyanov, R. P. Cherepanov, “Detection of Mechanisms for Unfair Increase of Scientometric Indexes,” *Computer tools in education*, no. 1, pp. 18–29, 2023 (in Russian); [doi:10.32603/2071-2340-2023-1-18-29](https://doi.org/10.32603/2071-2340-2023-1-18-29)

References

1. X. Bai, F. Xia, I. Lee, J. Zhang, and Z. Ning, “Identifying Anomalous Citations for Objective Evaluation of Scholarly Article Impact,” *PLOS ONE*, vol. 11, no. 9, p. e0162364, 2016; [doi:10.1371/journal.pone.0162364](https://doi.org/10.1371/journal.pone.0162364)
2. J. Chakraborty, D. Pradhan, H. S. Dutta, S. Nandi, and T. Chakraborty, “On Good and Bad Intentions behind Anomalous Citation Patterns among Journals in Computer Sciences,” in *arXiv*, 2018; [doi:10.48550/ARXIV.1807.10804](https://doi.org/10.48550/ARXIV.1807.10804)
3. T. Yu, G. Yu, Y. Song and M.-Y. Wang, “Toward the more effective identification of journals with anomalous self-citation,” *Malaysian Journal of Library & Information Science*, vol. 23, no. 2, pp. 25–46, 2018; [doi:10.22452/mjlis.vol23no2.2](https://doi.org/10.22452/mjlis.vol23no2.2)
4. S. Kojaku, G. Livan, and N. Masuda, “Detecting anomalous citation groups in journal networks,” *Scientific Reports*, vol. 11, p. 14524, 2021; [doi:10.1038/s41598-021-93572-3](https://doi.org/10.1038/s41598-021-93572-3)
5. J. Liu, F. Xia, X. Feng, J. Ren, and H. Liu, “Deep Graph Learning for Anomalous Citation Detection,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 6, pp. 2543–2557, 2022; [doi:10.1109/tnnls.2022.3145092](https://doi.org/10.1109/tnnls.2022.3145092)
6. J. Kemp, “Free public data file of 112+ million Crossref records,” in *Crossref.org*, 2020. [Online]. Available: <https://www.crossref.org/blog/free-public-data-file-of-112-million-crossref-records/>

7. C. Bron and J. Kerbosch, “Algorithm 457: finding all cliques of an undirected graph,” *Communications of the ACM*, vol. 16, no. 9, pp. 575–577, 1973; doi:10.1145/362342.362367
8. ORCID Inc., “About ORCID,” in *orcid.org*, 2023. [Online]. Available: <https://info.orcid.org/>
9. Digital Science & Research Solutions Inc., “The data in Dimensions – from idea to impact,” in *dimensions.ai*, 2023. [Online]. Available: <https://www.dimensions.ai/dimensions-data/>
10. D. Emelyanov, “Cliques In Science,” in *github.com*, 2022. [Online]. Available: https://github.com/Deimos-Apollon/Cliques_In_Science

Received 17-01-2023, the final version — 16-03-2023.

Dmitry Emelyanov, 3rd year Student of the bachelor’s degree program “Applied Mathematics and Informatics” of Saint Petersburg Electrotechnical University, ✉ dima31120251@gmail.com

Roman Cherepanov, 3rd year Student of the bachelor’s degree program “Applied Mathematics and Informatics” of Saint Petersburg Electrotechnical University, romacherepanov2002@gmail.com