

УДК 519.178, 004.9

ИССЛЕДОВАНИЕ БИБЛИОГРАФИЧЕСКИХ ССЫЛОК В MATH-NET.RU С ИСПОЛЬЗОВАНИЕМ ГРАФА ЦИТИРОВАНИЯ ЖУРНАЛОВ

А. А. Печников¹, [0000-0002-0683-0019], Д. Е. Чебуков², [0000-0001-9738-8707]

¹Институт прикладных математических исследований – обособленное подразделение ФИЦ «Карельский научный центр Российской академии наук», г. Петрозаводск

² Математический институт им. В.А. Стеклова Российской академии наук, г. Москва

¹pechnikov@krc.karelia.ru, ²tche@mi-ras.ru

Аннотация

По данным портала Math-Net.Ru построен граф цитирования журналов, дугами в котором являются цитирования с 2002 по 2021 годы. Для повышения достоверности модели был выбран временной интервал цитирования с 2010 по 2021 годы, когда распределение цитируемых статей стабилизировалось на уровне 3500–4500 цитирований за год. Исследована структура старения ссылок и показано, что время их полужизни равно 8 годам. Поэтому дата издания цитируемых статей была ограничена 2002 годом. Для построенного графа цитирования получены основные свойства, такие как маленький диаметр и высокая плотность, свидетельствующие о высоком уровне научных коммуникаций в Math-Net.Ru. Показано отсутствие «эффекта Матфея» как ярко выраженного преимущества в цитировании состоявшихся журналов по отношению к менее известным. Адекватность графа цитирования журналов Math-Net.Ru как модели научных коммуникаций подтверждается сравнением ранжирования журналов в графе цитирования с их рейтингом SCIENCE INDEX в eLIBRARY.RU. Показана прямая умеренная связь между двумя ранжированиями. Сделан ряд содержательных выводов, следующих из анализа графа цитирования.

Ключевые слова: библиографическая ссылка, граф цитирования журналов, старение библиографических ссылок, индекс Матфея, математический портал Math-Net.Ru.

ВВЕДЕНИЕ

Одной из первых публикаций, посвященных анализу сетей научного цитирования, является статья де Солла Прайс [1], вышедшая в 1965 году и, несомненно, связанная с появлением электронных хранилищ научных публикаций. Основные исследуемые объекты с тех пор и до настоящего времени – сети цитирования и сети соавторства, построенные на основе данных, получаемых из различных электронных источников, включая Google Scholar, которые исследуются с использованием математических моделей и дальнейшей содержательной интерпретацией результатов.

Идея определения значимости научных журналов посредством измерения их цитируемости возникла гораздо раньше, еще в 1927 году [2], когда еще не использовалась термины типа «центральность в теории графов и анализе сетей». Заметим, что графом чаще называют математический объект (граф в строгом смысле как множество вершин и множество пар вершин, к которому применим аппарат теории графов), а сетью – те же структуры с содержательным наполнением (как сети цитирования журналов или сети соавторства). В контексте данной статьи эти два понятия практически идентичны. В работе [3] сказано, что «... изучение сетей цитирования как окна в науку представляет собой проверенную временем традицию», и дан хороший обзор публикаций на английском языке. Здесь используемые термины «граф» (graph) и «сеть» (network) также понимаются в похожем смысле.

В зарубежных исследованиях, посвященных анализу научного цитирования, основой для получения исходных данных являются крупные и всем известные поисковые интернет-платформы, такие как Web of Science (WoS) или Scopus, имеющие инструменты типа Journal Citation Reports, и (не менее крупные) специализированные ресурсы, такие как PubMed (URL: <https://pubmed.ncbi.nlm.nih.gov>). Русскоязычных журналов и публикаций на этих ресурсах немного, например, в WoS количество русскоязычных публикаций составляет менее 0.5% от общего количества.

В российских исследованиях сетей цитирования журналов большое внимание уделяется экономическому направлению [4–6]. Представляется, что во многом это связано с использованием базы данных RePEc (Research Papers in

Economics – URL: <http://repec.org>), значительная часть материалов которой находится в свободном доступе. Но и здесь русскоязычные журналы занимают незначительное место.

Небольшое число публикаций, связанных с исследованиями библиографических цитирований в России, возможно, связано с определёнными сложностями в получении необходимых сведений из баз данных Научной электронной библиотеки eLIBRARY.RU (URL: <https://elibrary.ru>).

В отличие от указанных случаев авторы имеют доступ к сведениям Общероссийского математического портала Math-Net.Ru (URL: <http://www.mathnet.ru>), что в определенной степени позволяет исследовать указанную проблематику в области математических наук в России (см., например, [7]).

Math-Net.Ru – известный веб-ресурс, содержащий богатую коллекцию полнотекстовых архивов ведущих российских математических журналов и информацию об их авторах. К середине марта 2021 года на портале зарегистрировано 133578 авторов, 295964 научных статьи из 143 журнала (периодических изданий).

По классификации eLIBRARY.RU более 85% журналов Math-Net.Ru относятся к тематике «математика», а остальные – к тематикам «информатика», «кибернетика» и «автоматика и вычислительная техника» (несколько журналов относятся сразу к нескольким из этих тематик). По данным eLIBRARY.RU за 2018 год из первых пятидесяти журналов с наибольшим Science Index по тематике «математика» 46 журналов входят Math-Net.Ru.

Информационная система Math-Net.Ru индексирует списки литературы и хранит их в базе данных в структурированном виде [8]. Списки литературы всех публикаций объединены в одну таблицу базы данных, в которой в отдельных колонках хранятся данные об авторе, названии, годе, томе, страницах цитируемой публикации. В несколько упрощенном виде запись о цитировании выглядит следующим образом:

<цитирующий журнал><цитирующая статья><цитируемый журнал><цитируемая статья>.

Данные о статьях, в свою очередь, имеют поля, позволяющие точно идентифицировать их: список авторов, название, год, страницы и т. д. Самая старая статья, проиндексированная в Math-Net.Ru, опубликована в 1866 году.

Каждой индивидуальной ссылке соответствует одна запись в таблице. Такой подход облегчает задачу автоматического проставления гиперссылок на библиометрические базы данных, решает задачу поиска обратных ссылок, а также позволяет автоматически экспортировать ссылки в разные форматы: PDF, XML, HTML. Среди гиперссылок с элементов списков литературы есть и ссылки на статьи, индексированные в базе публикаций Math-Net.Ru. Таким способом осуществляется связь между цитирующей и цитируемой статьями.

Основная задача данного исследования заключается в построении графа цитирования журналов Math-Net.Ru и его исследовании. Показана адекватность модели путем сравнения ранжирований для графа цитирования, eLibrary.Ru и Web of Science.

1. ОПРЕДЕЛЕНИЕ ВРЕМЕННОГО ИНТЕРВАЛА

В информационной системе Math-Net.Ru данные накапливаются в течение последних 15 лет, начиная с 2006 года. Самые ранние цитирующие статьи (то есть имеющие библиографические ссылки на публикации Math-Net.Ru), относятся к 1936 году, а самые ранние цитируемые статьи датируются 1866 годом.

Наполнение информационной системы по годам по различным причинам было достаточно неоднородным. Для достоверности построения модели цитирования журналов на первом этапе исследований из общего количества 143 журналов были отобраны те журналы, которые полностью индексируются в Math-Net.Ru, а соответствующие выпуски обработаны и открыты на сайте (для проверки в ручном режиме в случае необходимости).

Анализ распределения статей по годам показывает, что режим стабилизации прибавления новых публикаций наступил к 2010 году, начиная с которого в Math-Net.Ru стало добавляться от 3500 до 4500 статей ежегодно. Поэтому на первом этапе была сделана выборка цитирований из базы Math-Net.Ru с 2010 по 2021 год для статей, опубликованных в этих 143 журналах.

Общее количество цитирований за этот период равно 162069, и сделаны они с 41381 статьи на 61022 статей. Если цитирования суммировать по журналам, то оказывается, что за указанный период 106 цитирующих журналов ссылаются на 133 цитируемых журнала. Учитывая пересечения множеств цитирующих и цитируемых журналов, получаем множество из 135 журналов.

На рисунке 1 приведена гистограмма распределения статей, имеющих исходящие ссылки, по годам. Понятно, что 2021 год находится в стадии наполнения, поэтому количество статей всего 175. По остальным 11 годам имеем примерно одинаковую картину.

Очевидно, что возможно самоцитирование журналов [9]. К примеру, журнал «Алгебра и анализ» за 2010–2021 годы в своих статьях имеет 2517 библиографических ссылки на другие статьи, из которых в 542 случаях цитируются статьи из этого же журнала. Самоцитирование журналов – сложное явление, например, в работе [10] утверждается, что самоцитирование журналов требует их исключения из рассмотрения в индексах цитирования. Мы не придерживаемся столь категоричной позиции, считая, что самоцитирование журналов является особой формой публикационной научной коммуникации, и поэтому учитываем его в наших моделях.

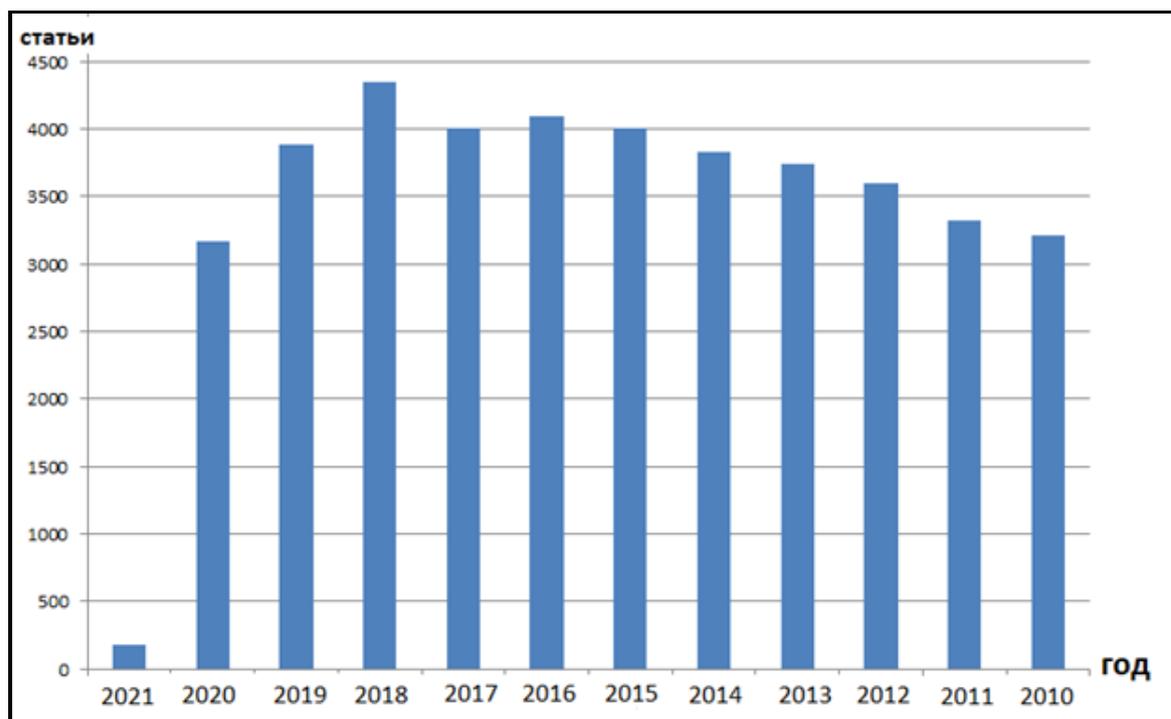


Рис. 1. Распределение цитирующих статей Math-Net.Ru в 2010–2021 годах

2. СТАРЕНИЕ ССЫЛОК

Для сформированного множества из 106 журналов, обязательно имеющих библиографические ссылки на другие журналы Math-Net.Ru за период 2010–2021 годов, была проанализирована возрастная структура ссылок.

Определим «возраст ссылки» как разницу между годом, когда опубликована статья, в которой приводится данная ссылка, и годом выпуска статьи, на которую ссылаются. «Время полужизни» цитируемости (или более точно «медиана хронологического распределения ссылок») определяется следующим образом: это такой момент времени, в который половина рассматриваемых ссылок относится к статьям, опубликованным позднее медианы, а другая половина – к статьям, более ранним, чем медиана.

На рисунке 2 показана интегральная возрастная структура исходящих ссылок данного множества журналов в виде графика, изображенного сплошной линией.

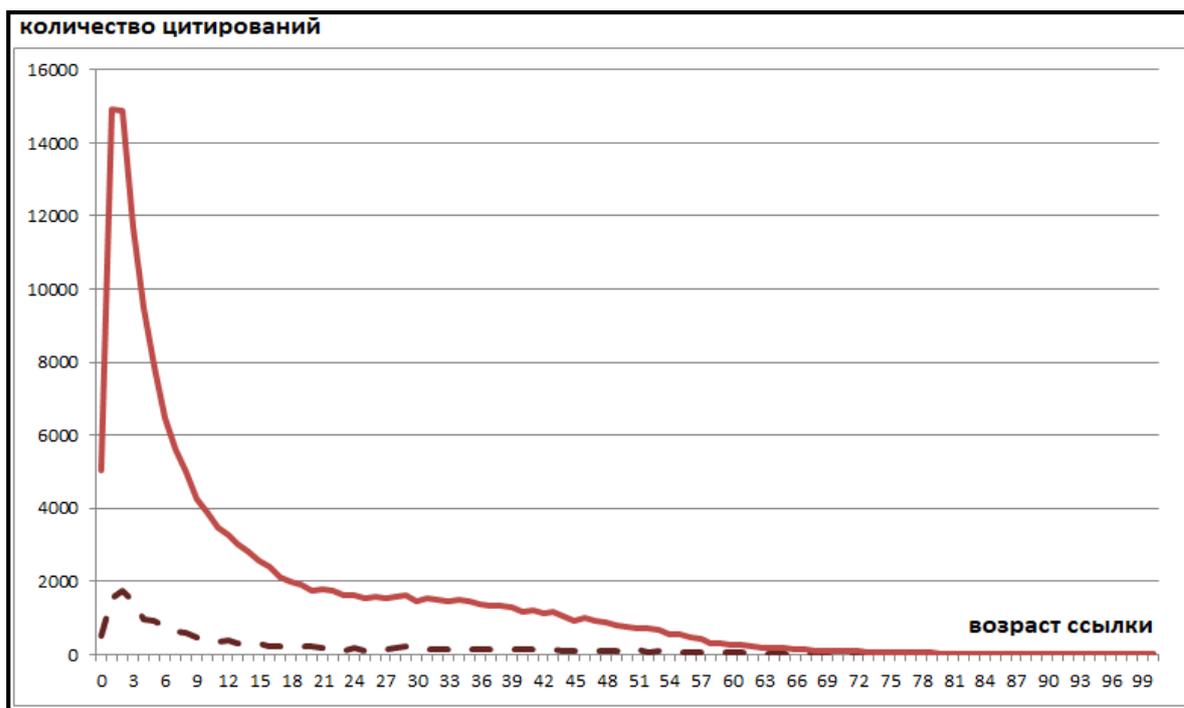


Рис. 2. Графики возрастной структуры исходящих ссылок

Например, значение в точке 0, равное 5060, вычисляется как суммарное количество всех ссылок, имеющих возраст, равный нулю. Возраст ссылки находится

в интервале от 0 до 152 лет. Максимальное количество цитирований (14924) приходится на статьи, опубликованные в предыдущем (по отношению к цитирующей публикации) году, то есть имеющие возраст цитирования, равный 1. Почти столько же цитирований приходится на возраст ссылки, равный 2, а затем происходит стремительный спад цитирований. Время полужизни ссылок для интегральной возрастной структуры равно 8 годам, то есть из 162 тысяч ссылок 81 тысяча имеет возраст не более 8 лет.

На этом же рисунке пунктирной линией изображен график возрастной структуры исходящих ссылок, сделанных в публикациях 2018 года. Значение в точке 0, равное 498, вычисляется как количество ссылок, сделанных в 2018 году на статьи, опубликованные в этом же году. Здесь максимум смещен на год правее, 1753 ссылки имеют возраст 2 года. Время полужизни ссылок 2018 года чуть меньше и равно 7,75 годам. Для сравнения, аналогичный график за 2010 год имеет практически такой же вид, за исключением того, что максимум достигается в точке 1, а время полужизни равно 9 годам.

3. СВОЙСТВА ГРАФА ЦИТИРОВАНИЯ

С учетом структуры старения и времени полужизни ссылок построим граф цитирования журналов Math-Net.Ru. Для этого будем использовать цитирования за 2010 по 2021 годы, ограничив годы издания цитируемых статей 2002 годом. Фактически мы используем найденный восьмилетний срок полужизни ссылок, и поэтому ссылки на статьи, опубликованные ранее 2002 года, не рассматриваются в нашей модели.

Указанное ограничение приводит к тому, что общее количество рассматриваемых журналов сократилось до 120, а количество цитирований – до 99 тысяч, причем почти 44 тысячи из них являются самоцитированиями. Уберем из множества журналов 13 журналов, имеющих только входящие или только исходящие ссылки, и сформированное множество журналов обозначим $J103$.

На множестве $J103$ построим граф цитирования журналов $G_{cit}(V, E, W)$, где:
 V – множество вершин (103 вершины, соответствующие журналам),
 E – множество дуг (3873 дуги, связывающие 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).

Сумма всех весов W – это и есть количество всех цитирований журналов из $J103$, равное 97109, причем 43910 из них являются самоцитированиями. У большей части журналов количество самоцитирований больше, чем количество цитирований наиболее цитируемого журнала.

Укажем основные характеристики графа $G_{cit}(V,E,W)$. По построению очевидно, что он является сильно связным: из любой вершины в любую другую можно перейти по пути с конечным числом дуг. Диаметр графа (максимальное количество дуг в таком пути) равно 4. Плотность графа (отношение числа дуг к максимально возможному) достаточно велика и составляет 0,369.

Интересна структура модулярности графа $G_{cit}(V,E,W)$. Напомним, что графы с высокой модулярностью имеют сильные связи между вершинами внутри модулей, но слабые связи между узлами в различных модулях. Если рассматривать дуги без учета их весов, то модулярность практически равна нулю. Это означает, что дуги в графе распределены достаточно равномерно.

В случае учета весов дуг получаем несколько иную картину. Значение коэффициента модулярности в этом случае равно $Q=0,356$. Мы здесь используем определение меры модулярности Q из [11]. Значение Q лежит в интервале $[-1,1]$, и разбиение считается хорошим, если значение Q больше 0,7.

Полученное значение Q не показывает сильной модулярности графа $G_{cit}(V,E,W)$, но говорит о некоторых тенденциях. Граф $G_{cit}(V,E,W)$ разбивается на 5 модулей, каждый из которых может быть содержательно интерпретирован по направлениям исследований: фундаментальная математика, математическое моделирование, экспериментальная и теоретическая физика, дискретная математика и прикладная математика и компьютерные науки. Перечислим для примера журналы модуля «экспериментальная и теоретическая физика»: «Компьютерная оптика», «Международный научно-исследовательский журнал», «Письма в Журнал экспериментальной и теоретической физики», «Нелинейная динамика», «Квантовая электроника», “Regular and Chaotic Dynamics”, «Теплофизика высоких температур» и «Успехи физических наук».

Поскольку значимость научных журналов характеризуется посредством их ранжирования на основе показателей, построенных с использованием данных о

цитировании статей, граф цитирования также может быть использован для этих целей. Значимость вершин для ориентированного графа можно определить различными способами, и каждый из них требует содержательной интерпретации. Оценка Page Rank (PR) дает возможность сравнить относительную «важность» вершин графа по аналогии с важностью веб-страниц в Вебе [12]. Содержательная интерпретация значимости вершин по PR в графе $G(V,E,W)$ может быть следующей: если представить себе некоего «учёного-сёрфера», бесконечное число раз перемещающегося от одного журнала к другому по пристатейным ссылкам, то с наибольшей вероятностью он будет попадать на журнал с наибольшим значением PR .

Для графа $G(V,E,W)$ были вычислены значения PR для каждой вершины с учетом петель и весов. Значения PR , упорядоченные по убыванию значений для пяти первых и пяти последних вершин, приведены в таблице 1.

Таблица 1. Значения PR для графа $G(V,E,W)$

Вершина	Название журнала	PR
Mzm	Математические заметки	0,0607
Sm	Математический сборник	0,0516
Ufn	Успехи физических наук	0,0422
zvmmf	Журнал вычислительной математики и математической физики	0,0412
Smj	Сибирский математический журнал	0,0408
...
Co	Компьютерная оптика	0,0015
cgtm	Contributions to Game Theory and Management	0,0015
vtamu	Вестник российских университетов. Математика	0,0015
svfu	Математические заметки СВФУ	0,0015
thsp	Theory of Stochastic Processes	0,0015

4. ЭФФЕКТ МАТФЕЯ

Одним из показателей репутации журналов является так называемый «эффект Матфея», введённый Мертоном [13]. По евангельской цитате «... всякому имущему дастся и приумножится, а у неимущего отнимется и то, что имеет» (Мф. 25:29) в отношении научного цитирования эффект трактуется как преимущество в

цитировании состоявшихся ученых по отношению к их менее известным коллегам. В статье [14] эффект Матфея, как отсутствие преимущества в цитировании состоявшихся журналов по отношению к менее известным, проверялся для журналов Math-Net.Ru. Исследование проводилось для множества, состоящего из 166 журналов, проиндексированных в Math-Net.Ru за весь период времени, начиная от самых ранних проиндексированных публикаций и заканчивая 2019 годом. Оказалось, что в значительной степени наблюдается эффект, обратный эффекту Матфея, то есть «богатые богатеют ...», хотя «бедные не беднеют».

Поскольку в данной работе мы исследуем множество из 103 журналов, сформированное по специальным условиям, и временной интервал сокращен до периода после 2001 года, было интересно проверить, как эффект Матфея выполняется в этом случае.

Индекс Матфея для выявления отклонений в цитировании, связанных с государственной принадлежностью авторов, впервые был введен в [15]. По аналогии определим MI для журналов следующим образом:

$$MI_i = (obs_i - exp_i) / exp_i,$$

где obs_i – реальный прирост числа цитирований, а exp_i – ожидаемый прирост число цитирований статей, опубликованных в журнале i за период времени $[t_0, t_1]$. Здесь t_0 и t_1 – начальный и конечный моменты времени, определяющие интервал вычисления индекса Матфея. Как правило, t_0 и t_1 – это начальный и конечный годы интервала исследований.

Мы опишем вычисление MI_i применительно к нашему случаю. Для каждого журнала в базе данных определяется общее количество статей, опубликованное в нем и суммарное количество их цитирований от самого начала заполнения базы до года t_0 , что позволяет вычислить среднюю цитируемость статьи в журнале на год t_0 . Затем в базе данных определяется прирост количества статей, опубликованных в журнале после года t_0 до года t_1 , который умножается на среднюю цитируемость. Это и есть exp_i для i -го журнала. Реальный прирост числа цитирований obs_i за период времени $[t_0, t_1]$ определяется из базы данных.

Если MI_i больше 0, то делается вывод, что журнал i за период $[t_0, t_1]$ получает больше цитирований, чем можно предположить, и наоборот. Особенность этого

подхода, используемая в нашем случае, заключается в том, что нет необходимости определять состоявшиеся и менее известные журналы посредством некоего ранжирования, которое в свою очередь надо обосновывать.

Приведем в качестве примера один из характерных случаев, когда были взяты $t_0=2018$, а $t_1=2020$. Часть результатов приведена в таблице 2.

Таблица 2. Значения основных параметров и индекса Матфея

<i>Название журнала</i>	<i>cit18</i>	<i>p18</i>	<i>mid18</i>	<i>cit20</i>	<i>p20</i>	<i>exp_i</i>	<i>obs_i</i>	<i>MI_i</i>
Вестник Санкт-Петербургского университета. Серия 10. Прикладная математика. Информатика. Процессы управления	5478	95	57,7	6026	127	1845,2	548	- 0,70
Квантовая электроника	74358	1421	52,3	97870	1725	15907,7	23512	0,48
Письма в Журнал экспериментальной и теоретической физики	58457	1489	39,3	75566	1885	15546,6	17109	0,10
Теплофизика высоких температур	21212	1027	20,7	23417	1116	1838,2	2205	0,20
Regular and Chaotic Dynamics	5038	278	18,1	6308	335	1032,9	1270	0,23
Компьютерные исследования и моделирование	2513	143	17,6	3186	200	1001,6	673	- 0,33
Системы и средства информатики	3234	191	16,9	4114	244	897,4	880	- 0,02
Программные системы: теория и приложения	1554	94	16,5	1775	111	281,0	221	- 0,21
Препринты Института прикладной математики им. М.В. Келдыша РАН	8458	579	14,6	10345	731	2220,4	1887	- 0,15
Математическое моделирование	6213	490	12,7	8186	590	1267,9	1973	0,56

Здесь $cit18$, $p18$ и $cit20$, $p20$ – количество цитирований журнала и количество опубликованных в нем статей за 2002–2018 и 2002–2020 годы соответственно, а $mid18$ – среднее количество ссылок на статьи в 2018 году.

Журналы упорядочены по убыванию параметра $mid18$, то есть «Вестник Санкт-Петербургского университета» на 2018 год является самым «богатым» по этому показателю (по общему количеству цитирований таковым является «Квантовая электроника»). Как видно из таблицы, для первой десятки «богатых» по среднему количеству цитирований пять журналов имеют положительное значение индекса Матфея, а пять – отрицательное. Примерно такая же картина наблюдается и в целом с той разницей, что количество положительных значений индекса составляет около 80%. Поэтому можно сказать, что ярко выраженного преимущества в цитировании состоявшихся журналов по отношению к менее известным не наблюдается.

5. СРАВНЕНИЕ РАНЖИРОВАНИЯ ЖУРНАЛОВ В MATH-NET.RU, ELIBRARY.RU И WEB OF SCIENCE

Было проведено сравнение результатов ранжирования журналов Math-Net.Ru по показателю PR с результатами ранжирования в eLIBRARY.RU и Web of Science (WoS). Для этого в eLIBRARY.RU был взят такой показатель, как место журнала в рейтинге SCIENCE INDEX (SI) [16]. Для расчета SI журналы приписываются к одному из 10 тематических направлений. Для исследования было взято направление «Mathematics, computer and information sciences», куда попали 58 журналов из J103. В таблице 3 показаны десять первых и пять последних журналов, упорядоченных по убыванию в рейтинге SI, их значения PR и порядковые номера по PR . Неожиданным выглядит только высокое место по рейтингу SI у «Информатики и автоматизации».

Таблица 3. Значения *PR* для журналов, входящих в SCIENCE INDEX

<i>Название журнала</i>	<i>SI</i>	<i>PR</i>	<i>#PR</i>
Успехи математических наук	1	0,0384	5
Математический сборник	2	0,0516	2
Известия Российской академии наук. Серия математическая	3	0,0322	7
Математические заметки	4	0,0607	1
Информатика и автоматизация	5	0,0027	47
Алгебра и анализ	6	0,0226	9
Журнал вычислительной математики и математической физики	7	0,0412	3
Труды Математического института имени В.А. Стеклова	8	0,0340	6
Сибирский математический журнал	9	0,0408	4
Функциональный анализ и его приложения	10	0,0181	14
.....		
Математические заметки СВФУ	54	0,0015	58
Математическая физика и компьютерное моделирование	55	0,0016	56
Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления	56	0,0026	49
Вестник Тверского государственного университета. Серия: Прикладная математика	57	0,0017	53
Дагестанские электронные математические известия	58	0,0017	52

В качестве математического инструмента был использован ранговый коэффициент Спирмена. В основу метода положен принцип нумерации значений статистического ряда. Каждому элементу совокупности присваивается порядковый номер в ряду, который будет упорядочен по уровню признака (например, по убыванию). Таким образом, ряд значений признака ранжируется, а номер каждого элемента будет его рангом. Пусть n – количество наблюдаемых значений признака, x_i – ранг i -го элемента первого статистического ряда, y_i – ранг i -го элемента второго статистического ряда, а d_i – разность между рангами. Тогда ранговый коэффициент Спирмена равен

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}.$$

Связь принято считать сильной, если $|\rho| \geq 0.7$, средней силы для $0.5 < |\rho| \leq 0.69$.

В нашем случае в качестве x_i были взяты места журналов в рейтинге *SI*, а в качестве y_i – места в рейтинге *PR*. Полученное значение $\rho=0,66$ при критическом значении 0,01 свидетельствует об умеренной прямой связи между двумя ранжированиями.

Таблица 4. Значения *PR* для журналов, входящих в WoS

Название журнала	#WoS	IF WoS	#PR	PR
Успехи физических наук	1	2,821	20	0,0422
Письма в Журнал экспериментальной и теоретической физики	2	1,399	17	0,0391
Успехи математических наук	3	1,345	16	0,0384
Regular and Chaotic Dynamics	4	1,285	7	0,0143
Квантовая электроника	5	1,184	8	0,0162
Известия Российской академии наук. Серия математическая	6	1,13	14	0,0322
Теплофизика высоких температур	7	1,085	2	0,0055
Теоретическая и математическая физика	8	0,854	12	0,0256
Алгебра и анализ	9	0,8	11	0,0226
Математический сборник	10	0,8	21	0,0516
.....			
Moscow Mathematical Journal	18	0,544	5	0,0081
Функциональный анализ и его приложения	19	0,487	9	0,0181
Теория вероятностей и ее применения	20	0,485	6	0,0130
Труды Математического института имени В. А. Стеклова	21	0,467	15	0,0340
Журнал математической физики, анализа, геометрии	22	0,227	1	0,0028

В WoS входят 22 журнала из *J103*. Была сделана выборка из WoS значений их импакт-факторов [17] за 2019 год и проведено сравнение с выборкой их значений *PR*. В таблице 4 показаны десять первых и пять последних журналов, упорядоченных по убыванию в рейтинге WoS, их значения *PR* и порядковые номера по *PR*.

Значение рангового коэффициента Спирмена, полученное в этом случае, равно -0,29 и показывает отсутствие корреляции.

В первой десятке заметно выделяется половина физических журналов. Но и удаление из этого списка журналов физического направления оставляет корреляцию практически такой же.

ЗАКЛЮЧЕНИЕ

В информационной системе Math-Net.Ru хранятся публикации, датированные, начиная с 1886 года. Однако в целях повышения достоверности построенной модели для исследований был выбран временной интервал с 2010 по 2021 годы. Анализ структуры старения ссылок для этих публикаций показал, что время полужизни исходящих цитирований равно 8 годам. Поэтому для построения графа цитирования журналов Math-Net.Ru были взяты цитирования в статьях за 2010 по 2021 годы с ограничением даты издания цитируемых статей 2002 годом.

Для построенного графа цитирования получены основные свойства, к которым прежде всего можно отнести высокую плотность графа и маленький диаметр, равный 4, что свидетельствует о высоком уровне научных коммуникаций как журналов Math-Net.Ru, так и ученых, публикующихся в них. Значимость вершин графа определена с помощью показателя Page Rank. Отсутствие «эффекта Матфея» свидетельствует о высокой репутации исследуемого множества журналов.

Адекватность графа цитирования журналов Math-Net.Ru как модели научных коммуникаций подтверждается сравнением ранжирования журналов (как вершин в графе цитирования по PR) с их рейтингом SCIENCE INDEX в eLIBRARY.RU. Показана прямая умеренная связь между двумя ранжированиями.

В то же время аналогичная проверка с ранжированием в WoS показывает полное отсутствие корреляции. И это также косвенно подтверждает адекватность модели: большая часть журналов Math-Net.Ru не входит в WoS, и их взаимные цитирования никак не влияют на импакт-фактор WoS. На импакт-фактор, естественно, влияют цитирования, сделанные с публикаций, входящих в WoS, но они в большинстве случаев не входят в Math-Net.Ru и, по-видимому, не являются российскими.

В.В. Налимов определял наукометрию как «... количественные методы изучения развития науки как информационного процесса» [18]. Считается, что сегодня наукометрия применяется для оценки научной результативности и финансирования различных научных единиц (институтов, команд, индивидуумов). Поскольку эта работа относится к области наукометрии, сформулируем один из выводов, основанных на проведенных исследованиях.

Один из недавних официальных документов [19] «... определяет способ расчета значения качественного показателя, характеризующего публикационную результативность научных организаций,» и разработан «... в целях методического обеспечения формирования государственного задания».

По этой методике журнал «... без квартиля», входящий в WoS, имеет «коэффициент качества» примерно в 8 раз больший, чем журнал «... из списка ВАК» (Высшей аттестационной комиссии). По нашей модели получается, что высокий рейтинг, полученный в результате оценки российскими коллегами, практически ничего не значит по сравнению с оценками зарубежных коллег, если журнал не входит в WoS. И даже если он входит в WoS, его оценка формируется без учета мнения российского математического сообщества. Это очень напоминает известное изречение ««Поищемъ сами в собѣ князя, иже бы володѣль нами и рядиль по ряду, по праву» [20]. Получается, что хотя нами и не «володеют», но «рядят по ряду».

СПИСОК ЛИТЕРАТУРЫ

1. *D.J. de Solla Price. Networks of scientific paper // Science. 1965. No. 149(3683). P. 510–515.*
2. *Gross P.L.K., Gross E.M. College Libraries and Chemical Education // Science. 1927. V. 66, Iss. 1713. P. 385–389.*
3. *Kas M., Carley K.M., Carley L.R. Trends in science networks: understanding structures and statistics of scientific networks // Social Network Analysis and Mining. 2012. No. 2. P. 169–187.*
4. *Алескеров Ф.Т., Бадгаева Д.Н., Писляков В.В., Стерлигов И.А., Швыдун С.В. Значимость основных российских и международных экономических журналов: сетевой анализ // Журнал новой экономической ассоциации. 2016. № 2. С. 193–205.*

5. Бредихин С.В., Ляпунов В.М., Щербакова Н.Г. Структура сети цитирования научных журналов // Проблемы информатики. 2017. № 2(35). С. 38–51.
6. Бредихин С.В., Ляпунов В.М., Щербакова Н.Г. Спектральный анализ сети цитирования научных журналов // Проблемы информатики. 2018. № 2(39). С. 24–40.
7. Печников А.А., Чебуков Д.Е. Исследование графа научного сотрудничества на основе цитирования в Math-Net.Ru // Информатизация образования и науки. 2020. № 3(47). С. 98–106.
8. Chebukov D., Izaak A., Misyurina O., Pupyrev Yu., Zhizhchenko A. Math-Net.Ru as a digital archive of the Russian mathematical knowledge from the XIX century to today // Lecture Notes in Comput. Sci. 2013. Vol. 7961. P. 344–348.
9. Писляков В.В. Самоцитирование и его влияние на оценку научной деятельности: обзор литературы. Часть I // Препринт.
URL: <https://arxiv.org/ftp/arxiv/papers/2109/2109.09192.pdf>
10. Heneberg P. From Excessive Journal Self-Cites to Citation Stacking: Analysis of Journal Self-Citation Kinetics in Search for Journals, Which Boost Their Scientometric Indicators // PLoS ONE. 2016. 11(4): e0153730.
<https://doi.org/10.1371/journal.pone.0153730>
11. Newman M.E., Girvan M. Finding and evaluating community structure in networks // Physical Review E. 2004. Vol. 69(2). P 026113.
12. Brin S., Page L. The Anatomy of a Large-Scale Hypertextual Web Search Engine // Computer Networks and ISDN Systems. 1998. No. 30. P. 107–117.
13. Merton R.K. The Matthew Effect in Science // Science. 1968. Vol. 159. Iss. 3810. P. 56–63.
14. Печников А.А. О репутационном статусе журналов Math-Net.Ru // Cloud of Science. 2020. Т. 7. № 4. С. 734–748.
15. Bonitz M., Bruckner E., Scharnhorst A. The Matthew Index – Concentration Patterns and Matthew Core Journals // Scientometrics. 1999. Vol. 44, No. 3. P. 361–378.
16. Методика расчета интегрального показателя научного журнала в рейтинге Science Index. URL: https://elibrary.ru/help_title_rating.asp

17. The Clarivate Analytics Impact Factor.
URL: <https://clarivate.com/webofsciencegroup/essays/impact-factor>
18. *Налимов В.В., Мультченко З.М.* Наукометрия: Изучение развития науки как информационного процесса. М: Наука, 1969. 191 с.
19. Методика расчета качественного показателя государственного задания «Комплексный балл публикационной результативности» для научных организаций, подведомственных Министерству науки и высшего образования Российской Федерации, на 2020 год.
URL: https://old.minobrnauki.gov.ru/common/upload/library/2020/09/main/Methodika_novaya.pdf
20. Повесть временных лет. Подготовка текста, перевод и комментарии О.В. Творогова. URL: <http://lib.pushkinskiydom.ru/?tabid=4869>
-

RESEARCH OF BIBLIOGRAPHIC REFERENCES IN MATH-NET.RU USING THE JOURNAL CITATION GRAPH

A. A. Pechnikov¹ [0000-0002-0683-0019], **D. E. Chebukov**² [0000-0001-9738-8707]

¹ *Institute of Applied Mathematical Research of the Karelian Research Centre of the Russian Academy of Sciences, Petrozavodsk*

² *Steklov Mathematical Institute of RAS, Moscow*

¹pechnikov@krc.karelia.ru, ²tche@mi-ras.ru

Abstract

According to the portal Math-Net.Ru a graph of journal citation is constructed. To increase the reliability of the model, the citation time interval was chosen from 2010 to 2021, when the distribution of citation articles stabilized at the level of 3500-4500 citations per year. The structure of link aging is studied and it is shown that their half-life is equal to 8 years. Therefore, the publication date of the cited articles was limited to 2002. For the constructed citation graph, the main properties, such as a small diameter and a high density, are obtained, indicating a high level of scientific communication in the Math-Net.Ru. The absence of the “Matthew effect” as a pronounced advantage in quoting established journals in relation to less well-known ones is shown.

Adequacy of the journal citation graph Math-Net.Ru as a model of scientific communication confirmed by comparing the ranking of journals in the citation graph with their SCIENCE INDEX rating in eLIBRARY.RU. A direct moderate relationship between the two rankings is shown. A number of meaningful conclusions are drawn from the analysis of the citation graph.

Keywords: *bibliographic reference, journal citation networks, aging of bibliographic references, Matthew index, mathematical portal Math-Net.Ru.*

REFERENCES

1. *D.J. de Solla Price.* Networks of scientific paper // *Science.* 1965. No. 149(3683). P. 510–515.
2. *Gross P.L.K., Gross E.M.* College Libraries and Chemical Education // *Science.* 1927. V. 66, Iss. 1713. P. 385–389.
3. *Kas M., Carley K.M., Carley L.R.* Trends in science networks: understanding structures and statistics of scientific networks // *Social Network Analysis and Mining.* 2012. No. 2. P. 169–187.
4. *Aleskerov F.T., Badgaeva D.N., Pislyakov V.V., Sterligov I.A., Shvydun S.V.* Znachimost' osnovnyh rossiiskih i mejdunarudnyh ekonomicheskikh jurnalov: setevoi analiz // *Journal of the New Economic Association.* 2016. № 2. P. 193–205.
5. *Bredikhin S.V., Lyapunov V.M., Shcherbakova N.G.* The structure of citation network of scientific journals // *Problems of Informatics.* 2017. №2 (35). P. 38–52.
6. *Bredikhin S.V., Lyapunov V.M., Shcherbakova N.G.* Spectral'nyi analiz seti citirovanija nauchnyh jurnalov // *Problems of Informatics.* 2018. № 2(39). P. 24–40.
7. *Pechnikov A., Chebukov D.* Issledovanie grafa nauchnogo sotrudnichestva na osnove citirovanija v Math-Net.Ru // *Informatizacija obrazovanija I nauki.* № 3(47). 2020. S. 98–106.
8. *Chebukov D., Izaak A., Misyurina O., Pupyrev Yu., Zhizhchenko A.* Math-Net.Ru as a digital archive of the Russian mathematical knowledge from the XIX century to today // *Lecture Notes in Comput. Sci.* 2013. Vol. 7961. P. 344–348.
9. *Pislyakov V.* Self-Citation and its Impact on Research Evaluation: Literature Review. Part I // Preprint.
URL: <https://arxiv.org/ftp/arxiv/papers/2109/2109.09192.pdf>.

10. *Heneberg P.* From Excessive Journal Self-Cites to Citation Stacking: Analysis of Journal Self-Citation Kinetics in Search for Journals, Which Boost Their Scientometric Indicators // PLoS ONE. 2016. 11(4): e0153730.

<https://doi.org/10.1371/journal.pone.0153730>

11. *Newman M.E., Girvan M.* Finding and evaluating community structure in networks // Physical Review E. 2004. Vol. 69(2). P 026113.

12. *Brin S., Page L.* The Anatomy of a Large-Scale Hypertextual Web Search Engine // Computer Networks and ISDN Systems. 1998. No. 30. P. 107–117.

13. *Merton R.K.* The Matthew Effect in Science // Science. 1968. Vol. 159. Iss. 3810. P. 56–63.

14. *Pechnikov A.A.* On the reputation status of Math-Net.Ru journals // Cloud of Science. 2020. Vol. 7. No. 4. P. 734–748.

15. *Bonitz M., Bruckner E., Scharnhorst A.* The Matthew Index – Concentration Patterns and Matthew Core Journals // Scientometrics. 1999. Vol. 44, No. 3. P. 361–378.

16. Metodika rascheta integral'nogo pokazatelja nauchnogo jurnala v reitinge Science Index. From URL: https://elibrary.ru/help_title_rating.asp (in Russian).

17. The Clarivate Analytics Impact Factor.
URL: <https://clarivate.com/webofsciencegroup/essays/impact-factor>

18. *Nalimov V., Mul'chenko Z.* Naukometrija: Izuchenie razvitija nauki kak informacionnogo processa. M.: Nauka, 1969. 191 s.

19. Metodika rascheta kachestvennogo pokazatelja gosudarstvennogo zadanija "Kompleksnyi ball publikacionnoi rezul'tativnosti" dlja nauchnyh organizacii, podvedomstvennyh Ministerstvu nauki i vys'shego obrazovanija Rossiiskoi Federacii, na 2020 god.

URL: https://old.minobrnauki.gov.ru/common/upload/library/2020/09/main/Metodika_novaya.pdf

20. Povest' vremennyh let. Podgotovka teksta, perevod i kommentarii O.V. Tvorogova. URL: <http://lib.pushkinskijdom.ru/?tabid=4869>

СВЕДЕНИЯ ОБ АВТОРАХ



ПЕЧНИКОВ Андрей Анатольевич – главный научный сотрудник Института прикладных математических исследований — обособленного подразделения ФИЦ «Карельский научный центр Российской академии наук», д. т. н. Сфера научных интересов – математическое моделирование, дискретная оптимизация, вебометрика, наукометрия.

Andrey Anatolievich PECHNIKOV – Chief Research Associate, Institute of Applied Mathematical Research of the Karelian Research Centre of the Russian Academy of Sciences, Doctor (DSc) of Technics. Research interests include mathematical modeling, discrete optimization, webometrics, scientometrics.

email: pechnikov@krc.karelia.ru,

ORCID: 0000-0002-0683-0019



ЧЕБУКОВ Дмитрий Евгеньевич – зав. информационно-издательским сектором Математического института им. В.А. Стеклова Российской академии наук, к. х. н. Сфера научных интересов – библиометрия, наукометрия, электронные библиотеки.

Dmitry Evgen'evich CHEBUKOV – Head of Information and Publishing Sector, Steklov Mathematical Institute of Russian Academy of Sciences, Candidate Chem. Sci. Research interests include bibliometrics, scientometrics, digital libraries.

email: tche@mi-ras.ru,

ORCID: 0000-0001-9738-8707

Материал поступил в редакцию 12 октября 2021 года