

БИБЛИОМЕТРИЧЕСКИЕ ИНДИКАТОРЫ В РЕСУРСАХ THOMSON REUTERS^{1, 2}



В. В. Писляков
Заместитель
директора
библиотеки
Национального
исследовательско-
го университета
«Высшая школа
экономики»,
кандидат физико-
математических
наук.

BIBLIOMETRIC INDICATORS IN THE THOMSON REUTERS INFORMATION RESOURCES

DOI 10.15826/B978-5-7996-1352-5.0005

This chapter examines bibliometric indicators related to citedness of journals, authors, research groups, institutions and whole countries. The introductory section deals with the basics of bibliometric analysis and features of citation databases. The author discusses the usage of various bibliometric indicators: the impact factor, average citedness, share of uncited papers, Eigenfactor and Article Influence Scores, Hirsch and Hirsch-type indices, and others. A special section investigates indicators of chronological distribution of references. Particular attention is paid to normalized indicators, including indicators normalized by research disciplines, as well as by publication sources. The final section emphasizes the importance of informed and reasonable use of bibliometric indicators in research policy-making, funding allocation, and faculty and research personnel recruitment.

В главе рассматриваются библиометрические индикаторы, оценивающие цитируемость журналов, авторов, научных коллективов, организаций и целых стран. Дается определение и обсуждается использование импакт-фактора и его вариаций, относительной цитируемости, ранговых метрик, коэффициента нецитируемости, «взвешенных» индикаторов (собственный фактор, индекс влияния статьи), индекса Хирша и ряда «Хирш-подобных» показателей и др. Специальный раздел посвящен показателям, характеризующим хронологическое распределение библиографических ссылок. Особое внимание уделено индикаторам нормализованной цитируемости, с нормализацией по областям науки и по журналам, а также их совокупному анализу. Глава содержит вводную часть, в которой излагаются основы библиометрического анализа и особенности баз данных научного цитирования. Финальный раздел подчеркивает необходимость грамотной и аккуратной трактовки библиометрических индикаторов при принятии административных решений, распределении грантов, осуществлении кадровой политики.

ВВЕДЕНИЕ

В данной главе мы рассмотрим ряд библиометрических индикаторов, которые либо со-

держатся в готовом виде, либо могут быть рассчитаны при помощи баз данных научного цитирования компании Thomson Reuters. Все

¹В данной главе использованы материалы издания: Писляков В. В. Библиометрические индикаторы: практикум. М.: НФПК, Инфра-М, 2014.

²Глава публикуется с сохранением авторской пунктуации.

эти индикаторы связаны с *цитируемостью* научных публикаций³.

Для каждого показателя в доступной форме будет представлен алгоритм его расчета и объяснен «физический смысл»: что именно он измеряет, когда может быть применен (а когда — нет), какие существуют потенциальные опасности и ограничения у предложенных методик.

Структура настоящей главы: в вводном разделе 2.1 рассказывается, что содержится в базах данных научного цитирования, каково наполнение и функции инструментов, предоставляемых Thomson Reuters, какие существуют обозначения, терминологические и методологические договоренности в литературе по библиометрии. Раздел 2.2 посвящен импакт-индикаторам, измеряющим среднюю цитируемость в расчете на одну статью; здесь же затрагивается вопрос о самоцитировании в научной литературе. В разделе 2.3 идет речь о подходах к кросс-дисциплинарному сравнению библиометрических характеристик — прежде всего, при помощи нормализованных индикаторов с различными способами нормализации, а также с использованием ранговых методов. В разделе 2.4 вводится понятие о показателях «экстремальной» (предельной) цитируемости, к которым отнесены доля высокоцитируемых статей и коэффициент нецитируемости. «Взвешенные» индикаторы, учитывающие при подсчете цитируемости научный уровень цитирующих журналов, освещаются в разделе 2.5. Индексу Хирша

и некоторым его модификациям посвящен раздел 2.6. В разделе 2.7 исследуются показатели, характеризующие хронологическую структуру распределения ссылок и, наконец, Заключение завершает данную главу.

2.1. Библиометрические инструменты. База данных Web of Science Core Collection и аналитические надстройки

Начнем мы с описания инструмента Web of Science Core Collection (Web of Science CC), продукта компании Thomson Reuters. Web of Science CC относится к «библиометрическим базам данных» или, как еще говорят, «базам данных (научного) цитирования» (*citation database, citation index*). Web of Science CC размещается на более широкой платформе Web of Science, включающей, помимо Web of Science CC, специализированные базы данных по патентам, ряд тематических реферативных баз данных, национальные (страновые) базы научного цитирования и др.

2.1.1. Принципы организации библиометрических баз данных. Методологические замечания

Из чего состоят базы научного цитирования, и что именно можно в них найти? Основной контент библиометрических баз данных — научные журналы. При этом в базе данных *не содержится полного текста статей* этих журналов. О каждой статье хранится (и может быть выдана пользователю), как правило, следующая информация:

³Необходимо заметить, что английское «citation», как справедливо указывалось, например, А. В. Поляевым или Р. С. Гиляревским, означает не «цитирование» или тем более «цитату», а лишь библиографическую ссылку. При строгом подходе следует говорить не о цитируемости, а о числе полученных ссылок, «citation index» переводить как «указатель ссылок» и т. д. Тем не менее в силу ограниченности языка, необходимости использовать синонимы и трудноискренимой практики, уже сложившейся в русскоязычной литературе, в рамках настоящей главы «цитирование» будет эквивалентом «ссылке». При этом, конечно, не подразумевается, что если одна статья «цитирует» другую, то в первой содержится выдержка, фрагмент из второй (то, что обозначается английским словом «quotation»). Речь всего лишь о том, что вторая статья фигурирует в списке использованной литературы первой статьи.

— библиографические сведения о статье («выходные данные»: автор(ы), название статьи, название журнала, год выхода, том, номер, страницы);

— аннотация статьи (реферат) — в том случае если она имелась в исходном тексте публикации; базы цитирования не составляют аннотации тех произведений, в которых они изначально отсутствуют;

— ключевые слова; иногда это несколько наборов ключевых слов — например, слова, приписанные статье в оригинальном

тексте публикации («авторские ключевые слова»); ключевые слова, «назначенные» базой данных цитирования исходя из ее внутреннего тезауруса ключевых слов и автоматических алгоритмов; или ключевые слова, заимствованные из тематических реферативных баз данных, где есть та же самая статья;

— тематика (рубрика), приписанная статье, и тип публикации (см. далее);

— организации, в которых работают авторы, т. н. «аффилиации», с почтовыми

UvrD facilitates DNA repair by pulling RNA polymerase backwards

By: Epshtein, V (Epshtein, Vitaly)^[1]; Kamarthapu, V (Kamarthapu, Venu)^[1,2]; McGary, K (McGary, Katelyn)^[1]; Svetlov, V (Svetlov, Vladimir)^[1]; Ueberheide, B (Ueberheide, Beatrix)^[1]; Proshkin, S (Proshkin, Sergey)^[3]; Mironov, A (Mironov, Alexander)^[3,4]; Nudler, E (Nudler, Evgeny)^[1,2]

NATURE

Volume: 505 Issue: 7483 Pages: 372+

DOI: 10.1038/nature12928

Published: JAN 16 2014

[View Journal Information](#)

Abstract

UvrD helicase is required for nucleotide excision repair, although its role in this process is not well defined. Here we show that *Escherichia coli* UvrD binds RNA polymerase during transcription elongation and, using its helicase/translocase activity, forces RNA polymerase to slide backward along DNA. By inducing backtracking, UvrD exposes DNA lesions shielded by blocked RNA polymerase, allowing nucleotide excision repair enzymes to gain access to sites of damage. Our results establish UvrD as a bona fide transcription elongation factor that contributes to genomic integrity by resolving conflicts between transcription and DNA repair complexes. Furthermore, we show that the elongation factor NusA cooperates with UvrD in coupling transcription to DNA repair by promoting backtracking and recruiting nucleotide excision repair enzymes to exposed lesions. Because backtracking is a shared feature of all cellular RNA polymerases, we propose that this mechanism enables RNA polymerases to function as global DNA damage scanners in bacteria and eukaryotes.

Keywords

KeyWords Plus: NUCLEOTIDE EXCISION-REPAIR; ESCHERICHIA-COLI-UVRD; TRANSCRIPTION ELONGATION; HELICASE-II; GENOME INSTABILITY; GENE-PRODUCT; IN-VITRO; E. COLI; DAMAGE; PROTEIN

Author Information

Reprint Address: Nudler, E (reprint author)

+ NYU, Sch Med, Dept Biochem & Mol Pharmacol, New York, NY 10016 USA.

Addresses:

+ [1] NYU, Sch Med, Dept Biochem & Mol Pharmacol, New York, NY 10016 USA

+ [2] NYU, Sch Med, Howard Hughes Med Inst, New York, NY 10016 USA

[3] State Res Inst Genet & Select Ind Microorganisms, Moscow 117545, Russia

+ [4] Russian Acad Sci, Engelhardt Inst Mol Biol, Moscow 119991, Russia

E-mail Addresses: evgeny.nudler@nyumc.org

Рис. 7. Запись в библиометрической базе данных на статью из журнала Nature. Скриншот интерфейса базы данных Web of Science CC

адресами мест работы и, иногда, электронными адресами авторов;



– список цитируемой в статье литературы — именно это поле является ключевым для базы данных и делает ее «библиометрической» базой данных;


– различные второстепенные поля: номер ISSN журнала, язык оригинального документа, название и адрес издательства и др.


Таким образом, если журнал «расписывается» (индексируется) библиометрической базой данных, это означает, что о каждой публикации этого журнала мы сможем узнать из базы данных сведения, перечисленные выше. Полного текста статей при этом в базе не будет. В качестве иллюстрации на рис. 7 приведен пример библиографической записи Web of Science CC на некую журнальную статью, а на рис. 8 — список


Cited References: 54
(from Web of Science Core Collection)


From: UvrD facilitates DNA repair by pulling RNA polymerase backwards ...[More](#)

Select Page |   | [Save to EndNote online](#) ▾ | [Add to Marked List](#)

1. **A physical interaction of UvrD with nucleotide excision repair protein UvrB**
By: Ahn, B
MOLECULES AND CELLS Volume: 10 Issue: 5 Pages: 592-597 Published: OCT 31 2000
 [View Abstract](#)

2. **TRANSCRIPTIONAL CONTROL OF THE UVRD GENE OF ESCHERICHIA-COLI**
By: ARTHUR, HM; EASTLAKE, PB
GENE Volume: 25 Issue: 2-3 Pages: 309-316 Published: 1983


3. **A ratchet mechanism of transcription elongation and its control**
By: Bar-Nahum, G; Epshtein, V; Ruckenstein, AE; et al.
CELL Volume: 120 Issue: 2 Pages: 183-193 Published: JAN 28 2005
 [View Abstract](#)

4. **Damage recognition in nucleotide excision repair of DNA**
By: Batty, DP; Wood, RD
GENE Volume: 241 Issue: 2 Pages: 193-204 Published: JAN 11 2000
 [View Abstract](#)


5. **TRANSCRIPT CLEAVAGE FACTORS FROM ESCHERICHIA-COLI**
By: BORUKHOV, S; SAGITOV, V; GOLDFARB, A
CELL Volume: 72 Issue: 3 Pages: 459-466 Published: FEB 12 1993
 [View Abstract](#)

Рис. 8. Список цитируемой в статье литературы.
Скриншот интерфейса базы данных Web of Science CC

литературы этой статьи (он открывается при нажатии соответствующей ссылки).

Помимо журнального контента, библиометрические базы данных могут включать в себя также некоторое количество трудов конференций и, с недавнего времени, книг. При этом поля, попадающие в базу данных, аналогичны тем, которые перечислены для журналов. Только в случае трудов конференций структурной единицей будет не статья, а доклад на конференции; в случае книг — глава из книги. В случае сборников трудов конференций речь идет именно о сборниках текстов докладов, а не о кратких тезисах. Обычно текст такого доклада занимает от 4 страниц и более и в обязательном порядке имеет список использованной литературы. За исключением своего предназначения, он ничем не отличается от статьи в журнале (и впоследствии часто становится статьей). В серьезных научных монографиях также имеется список литературы по главам. Это дает возможность базам данных трактовать доклады и главы в полном подобии статьям и индексировать эти документы на тех же принципах (но также без размещения полного текста).

В базах научного цитирования каждый журнал имеет «тематическую привязку», он отнесен к той или иной научной дисциплине (может быть отнесен более чем к одной). При этом при расчете практически всех индикаторов, о которых пойдет речь в настоящей главе, тематическая рубрика *статьи* определяется в базе данных цитирования на основании тематической рубрики *журнала*, в котором она опубликована. Все статьи одного журнала имеют одну и ту же рубрику (рубрики). Мы будем говорить, что статья/журнал относятся к той или иной дисциплине или научной

области. Два последних понятия используются нами как синонимы.

Помимо дисциплины, каждой публикации в библиометрических базах данных присваивается *тип документа*. Это может быть научная статья (Article), научный обзор (Review), заметка редактора (Editorial), письмо (Letter), книжная рецензия (Book Review) и др. При расчете ряда библиометрических индикаторов могут учитываться не все, а лишь некоторые типы публикаций — чаще всего это Article и Review. Кроме того, замечено, что разные типы публикаций обычно получают разное число ссылок. Так, документы типа Review, «научный обзор», в среднем цитируются более активно, чем типа Article, «научная статья». Тем не менее словом «статья» мы будем обозначать, как часто это делается в обыденной практике, любую публикацию в журнале. В тех разделах, где будет важно отличать «публикацию» от «статьи» (т. е. от публикации типа «научная статья»), мы будем прибегать к английскому «Article» — так избежим двусмысленности.

В библиометрии статью считают публикацией той или иной лаборатории/организации/страны на основании институциональных адресов, которые указаны авторами в статье и, соответственно, перенесены в поле «address» (или «affiliation» и т. д.) базы данных. В данном случае при анализе не интересуются деталями авторской биографии — как он менял места работы, одновременно совмещал их и т. д. «То, что указано в статье» — это окончательный ответ на вопрос, в какой организации работал автор, ее написавший. При этом автор, конечно, может указать в статье сразу несколько мест работы — и в таком виде эта информация будет перенесена в базы цитирования.

Про отнесение публикаций к авторам, организациям и странам следует пояснить еще один момент, связанный с соавторством. Существует несколько способов счета статей, написанных несколькими авторами. Самый простой из них — «полный счет», whole (или total) counting, который предполагает, что каждому из соавторов засчитывается по одной статье. Написал автор статью в одиночку или в сотрудничестве с 20 коллегами — это все равно плюс одна публикация в его статистике. То же самое для организаций и стран: если организация (страна) хотя бы один раз указана в списке мест работы авторов статьи, эта публикация считается как целиком принадлежащая ей (и всем остальным организациям/странам соавторов наравне).

Помимо этого есть несколько способов «дробного счета», fractional counting. Если авторов в статье n , то считается, что каждый написал $1/n$ статьи, и сложением именно таких долей находится суммарная публикационная активность автора по всем его работам. То же самое для организаций: каждая статья распределяется по всем организациям в равных долях или в долях, пропорциональных числу авторов из соответствующей организации. Дробный счет применяется как для публикаций, так и для цитирований, которые они получили.

Во всех индикаторах, которые будут рассмотрены в настоящей главе, будет использоваться только метод whole counting. Его отличает простота, однако следует помнить, что он не обеспечивает аддитивность большинства показателей. Например, отвечая на вопрос «сколько всего статей написали авторы организации», нельзя приводить арифметическую сумму числа статей, написанных каждым автором этой организации. Если два или более ученых

из данного института написали статью в соавторстве, то каждый из них получит по одной статье в свою публикационную статистику. Таким образом, простая сумма по индивидуальным показателям авторов даст двойку или больше, в то время как для статистики организации как целого это, очевидно, одна статья. Кроме того, даже если автор один, но он указал, например, два места работы в двух разных странах (это вполне возможно), то такая публикация при использовании whole counting одновременно поднимет на единицу статистику двух разных лабораторий, двух разных организаций и двух разных стран.

2.1.2. Web of Science Core Collection

Итак, Web of Science CC является базой данных цитирования, «библиометрической» базой данных. Она содержит семь блоков (или «изданий», editions) и в сумме охватывает более 12 000 научных журналов, 12 000 серий трудов конференций и более 50 000 книг. Web of Science CC является ядром библиометрических продуктов Thomson Reuters, их главной частью. Блоки, из которых состоит Web of Science CC, содержат различные типы документов по различным разделам науки (в самом широком смысле этого слова):

- Science Citation Index Expanded (SCIE) — журналы по естественным, техническим, медицинским наукам (физика, химия, математика, биология, информатика и вычислительная техника, медицина и т. д.); архив с 1898 г.
- Social Sciences Citation Index (SSCI) — журналы по общественным наукам (экономика, менеджмент, социология, право, политология, история и т. д.); архив с 1898 г.
- Arts & Humanities Citation Index (A&HCI) — журналы по гуманитарным наукам (исто-

рия, литературоведение, искусствоведение, религиоведение и т. д.); архив с 1975 г.

- Conference Proceedings Citation Index – Science – труды конференций, посвященных естественным, техническим, медицинским наукам; архив с 1990 г.
- Conference Proceedings Citation Index – Social Science & Humanities – труды конференций, посвященных общественным и гуманитарным наукам; архив с 1990 г.
- Book Citation Index – Science – книги, посвященные естественным, техническим, медицинским наукам; архив с 2005 г.
- Book Citation Index – Social Sciences & Humanities – книги, посвященные общественным и гуманитарным наукам; архив с 2005 г.

В журнальных блоках есть издания, дублирующиеся в разных индексах, – речь идет о мультидисциплинарных журналах, относящихся к различным типам наук. Например, журнал *Scientometrics* отнесен к рубрике «Information Science & Library Science» в SSCI и к «Computer Science Multidisciplinary Applications» в SCIE.

В настоящее время Web of Science CC индексирует около 150 российских изданий в SCIE, 3 журнала в SSCI и 5 – в A&HCI.

2.1.3. Аналитические надстройки: Journal Citation Reports, Essential Science Indicators, InCites

Повторим, что для каждой публикации в Web of Science CC можно узнать, какие документы она цитирует и, наоборот, какие документы цитируют ее. Следовательно, можно собирать статистику цитируемости. Для комплексных задач сделать это нелегко, поэтому Web of Science CC дополняют аналитические инструменты, в частности базы данных Journal Citation Reports (JCR), Essential Science Indicators (ESI) и InCites. Все

эти продукты используют Web of Science CC как свою основу и лишь собирают, агрегируют из Web of Science CC данные по различным «информационным единицам» (журналам, ученым, организациям, странам и т. д.).

JCR – база данных по библиометрическим показателям *журналов как целого*. Публикуются данные о количестве выходящих в журнале статей, числе полученных журналом ссылок, хронологическом распределении сделанных/полученных ссылок, импакт-факторе журнала (см. далее п. 2.2.1) и др.

ESI – база данных по библиометрическим индикаторам авторов, организаций, стран, журналов. Публикуются данные о количестве вышедших у автора/организации/страны/журнала статей и их цитируемости (по журналам информация в ESI менее подробная, чем в JCR). Ограничения: в ESI включены только те организации и авторы, которые попали в 1 % наиболее цитируемых хотя бы в одной научной дисциплине; и только те журналы и страны, которые попали в 50 % наиболее цитируемых хотя бы в одной научной дисциплине. Помимо этого в ESI имеется специальный раздел по мировым высокоцитируемым статьям (попавшим в 1 % самых цитируемых среди тех, которые вышли в заданном году в заданной научной области, см. далее п. 2.4.1), среднемировым библиометрическим показателям («baselines») и перспективным научным «фронтам», которые определяются специальной библиометрической процедурой (об этом подробнее см. на с. 224).

InCites – аналитический инструмент, позволяющий проводить детальный и глубокий анализ библиометрических показателей организаций, отдельных ученых, стран. Единственный из перечисленных продуктов использует нормализацию цитируемости по областям науки и по журналам. Во

второй половине 2014 г. вышла новая, расширенная версия продукта и, поскольку изменения коснулись не только интерфейса, но и методологической основы используемых индикаторов, мы будем обозначать две версии как InCites1 и InCites2.

Подчеркнем еще раз, что все данные, приведенные в JCR, ESI и InCites, рассчитаны по Web of Science CC и, строго говоря, содержатся в Web of Science CC. Однако вручную их крайне сложно оттуда извлечь, поэтому готовую агрегированную информацию по журналам, авторам, организациям и странам приводят для удобного использования в специальных аналитических «надстройках» JCR, ESI и InCites.

2.2. Индикаторы влиятельности статей (импакт-индикаторы)

К этому наиболее распространенному классу индикаторов цитируемости относятся показатели, оценивающие число ссылок, полученных в среднем одной статьей, входящей в некоторое заданное множество публикаций. Это могут быть статьи из какого-либо журнала или работы определенного автора, исследовательского коллектива, организации, целой страны и т. д.

Чаще всего исследуется множество статей, опубликованных за определенный фиксированный промежуток времени. Например, это могут быть статьи, написанные в некотором исследуемом университете U за пятилетний промежуток 2009–2013 гг. Этот интервал, в течение которого выходили оцениваемые статьи, называется «публикационным окном» (publication window). Не менее важен временной интервал, в течение которого *выходили те статьи, ссылки из которых мы учитываем* при подсчете импакт-индикатора. Например, это могут быть статьи, вышедшие в 2013 г. Этот промежуток

времени называется «окном цитирования» (citation window).

При подсчете импакт-индикаторов (и большинства библиометрических индикаторов в принципе) необходимо четкое обозначение как публикационного окна, так и окна цитирования, — в противном случае методика, определяющая измеряемый индикатор, будет неполна, а корректный и однозначный подсчет показателя будет невозможен. Итак, в приведенном нами примере мы бы считали среднее число ссылок, присутствующих в статьях 2013 г. выхода, на работы, опубликованные в 2009–2013 гг. учеными, работающими в вузе U . «Среднее» обозначает здесь усреднение по числу оцениваемых статей: найденное абсолютное число ссылок должно быть поделено на число статей в публикационном окне (т. е. на число статей, опубликованных учеными вуза U в 2009–2013 гг.).

Рассмотрение импакт-индикаторов мы начнем с журнальных показателей, поскольку впервые они были введены именно для журналов и на журнальной литературе легче всего проиллюстрировать их закономерности.

2.2.1. Импакт-фактор журнала

Как сказано выше, импакт-индикаторы являются наиболее распространенными среди показателей цитируемости. А самым известным и широко используемым среди них является импакт-фактор журнала (journal impact factor). При его подсчете используется двухлетнее публикационное окно и однолетнее окно цитирования.

Импакт-фактор журнала изменяется из года в год, поэтому рассчитывается для конкретного (отчетного) года. Для года Y импакт-фактор журнала равен отношению числа всех ссылок, полученных в году

У статьями данного журнала, вышедшими в годах $Y-1$ и $Y-2$, к числу этих статей (т. е. к числу статей журнала, вышедших в годах $Y-1$ и $Y-2$). Таким образом, публикационное окно — два года [$Y-1$; $Y-2$] (оценивается средняя цитируемость этого множества статей журнала), окно цитирования — один год [Y] (учитываются цитирования, сделанные в этом году).

Другими словами, импакт-фактор характеризует среднее число ссылок, полученных в отчетном году статьями журнала, опубликованными в течение двух предыдущих лет. Проиллюстрируем это конкретным примером. В 2013 г. во всех журналах, охваченных Web of Science CC, было сделано 125 ссылок на статьи журнала «Успехи химии», опубликованные в 2012 г., и 154 ссылки на его статьи, опубликованные в 2011 г. Итого статьи «Успехов», вышедшие в 2011–2012 гг., получили в 2013 г. $125 + 154 = 279$ цитирований. При этом в 2012 г. журнал опубликовал 54 статьи, в 2011 г. тоже 54, итого за промежуток 2011–2012 гг. (публикационное окно)

вышло 108 статей. Таким образом, импакт-фактор журнала «Успехи химии» в 2013 г. равен отношению $279 : 108 = 2,58$.

Все эти данные (как итоговый импакт-фактор, так и слагаемые его числителя и знаменателя) можно найти в базе данных JCR. Там они публикуются ежегодно, данные за очередной год обычно появляются в июне-июле следующего года (т. е. импакт-факторы 2014 г. можно узнать летом 2015 г. и т. д.). На рис. 9 приведен скриншот базы данных JCR с журналами по акустике, которые упорядочены по убыванию импакт-фактора. Кроме того, некоторые издательства размещают текущие значения импакт-факторов своих журналов на их веб-страницах, в открытом доступе. Нередко они сопровождаются ранговыми показателями (см. далее, п. 2.3.4).

На данный момент (конец 2014 г., последнее вышедшее издание — JCR-2013) максимальный импакт-фактор имеет журнал CA — A Cancer Journal for Clinicians, $IF = 162,5$. Это совершенно невероятный показатель (журнал публикует малое число статей, но получает много ссылок), и ближайший «конкурент», журнал New England Journal of Medicine, отстает от него ровно в три раза, $IF = 54,4$. Как и многие другие распределения в библиометрии, распределение журналов по импакт-фактору сильно сдвинуто относительно нормального: существует несколько журналов-лидеров и обширное множество изданий с низким импактом. В итоге медиана распределения импакт-факторов в журналах, входящих в SCIE (т. е. журналов по естественным, техническим, медицинским наукам), равна 1,41. То есть среди всех 8474 журналов половина имеет импакт-фактор выше 1,41, а половина — ниже.

Обратим внимание: JCR не публикует импакт-факторы гуманитарных журналов (т. е.

	Full Journal Title	Total Cites	Journal Impact Factor ▼
1	ULTRASCHALL IN DER MEDIZIN	1.332	4.645
2	JOURNAL OF VIBRATION AND CONTROL	3.288	4.355
3	ULTRASONICS SONOCHEMISTRY	5.796	3.816
4	ULTRASOUND IN OBSTETRICS & GYNECOLOGY	8.105	3.140
5	IEEE Transactions on Audio Speech and Language Processing	3.492	2.625
6	ULTRASOUND IN MEDICINE AND BIOLOGY	7.532	2.099
7	JOURNAL OF SOUND AND VIBRATION	21.330	1.857

Рис. 9. Журналы по акустике, в порядке убывания импакт-фактора. Скриншот интерфейса базы данных JCR (Thomson Reuters)

тех, которые входят *только* в базу A&HCI). Факторов, обуславливающих это, несколько. Прежде всего, в гуманитарных науках низкий уровень цитируемости статей, в том числе из-за того, что большую роль играет книжная литература, ее относительная роль выше, чем в общественных и, тем более, естественных/технических/медицинских науках. Кроме того, в большей мере цитируются «старые» материалы, опубликованные вне рамок двухлетнего окна цитирования, используемого при подсчете импакт-фактора. Все это делает практически невозможным сравнение научного уровня гуманитарных журналов при помощи импакта. Чтобы не вводить ученых и издателей в заблуждение, Thomson Reuters следует политике вообще не публиковать импакт-факторы для изданий по гуманитарным наукам.

2.2.2. Пятилетний импакт-фактор журнала, индекс оперативности

Когда говорят о «классическом» импакт-факторе, или «гарфилдовском» импакт-факторе, или просто об импакт-факторе без уточнений, то имеют в виду показатель, введенный в п. 2.2.1. Он учитывает цитирования, которые статьи журнала получают в течение двух лет после своего выхода. Однако в ряде научных дисциплин, особенно в социальных и гуманитарных науках, профессиональное сообщество не успевает в полной мере воспринять новое знание за столь короткий срок, как два года, и целесообразно использовать показатель с более широким публикационным окном. Поэтому в базе данных JCR также публикуются значения пятилетнего импакт-фактора журналов.

Для года Y пятилетний импакт-фактор журнала равен отношению числа всех ссылок, полученных в году Y его статьями, вышедшими в годах с $Y-1$ по $Y-5$, к числу этих

статей (т. е. к числу статей журнала, вышедших в годах с $Y-1$ по $Y-5$). Таким образом, импакт-фактор характеризует среднее число ссылок, сделанных в отчетном году на статьи журнала, опубликованные в течение пяти предыдущих лет.

Если упорядочить журналы по убыванию пятилетнего импакт-фактора, результат будет отличаться от ранжирования по обычному импакту: журналы, чьи материалы устаревают медленнее, т. е. те издания, на статьи которых делается значительное число ссылок даже через пять лет после их публикации, будут иметь преимущество перед журналами, получающими основное число ссылок на материалы менее чем трехлетней давности. Известно, что скорость «старения» публикуемого журналом знания в первую очередь зависит от его дисциплинарной области (подробнее см. далее, раздел 2.7).

Максимальный пятилетний импакт-фактор для журналов из SCIE равен 107,7, медиана — 1,50. Она несколько выше медианы двухлетнего импакт-фактора (1,41), это свидетельствует о том, что пятилетний показатель более полно охватывает цитирования, полученные журналами.

Заметим теперь, что при вычислении как импакт-фактора, так и пятилетнего импакт-фактора не учитываются ссылки, сделанные на те статьи журнала, которые вышли непосредственно в отчетном году. Они «пропадают» для журнала, если мы используем только эти два показателя. Тем не менее такие цитирования встречаются и, более того, их число имеет сейчас тенденцию к увеличению по причине все более частого размещения препринтов в Интернете, открытия специальных разделов на сайтах издательств, где публикуются предварительные версии статей, принятых в печать, а также общего ускорения производствен-

ного цикла научных издательств. Показатель, фиксирующий цитирования «того же года», также публикуется в базе данных JCR и называется *immediacy index*. Будем называть его «индексом оперативности» (другой вариант перевода, также предлагавшийся в литературе, — «индекс немедленного цитирования»).

Индекс оперативности предполагает одностороннее публикационное окно и одностороннее окно цитирования, причем они совпадают — это отчетный год Y . Для вычисления индекса необходимо разделить число всех ссылок, полученных в году Y статьями журнала, вышедшими в том же году Y , на число этих статей. Индекс оперативности показывает, насколько быстро ученый мир реагирует на статьи журнала, как скоро воспринимает его тексты и использует их при воспроизводстве научного знания.

Самый большой индекс оперативности в JCR-2013 (у того же журнала *CA — A Cancer Journal for Clinicians*) — 27,76. Медиана по журналам из SCIE предсказуемо мала — 0,27. Практически для каждого журнала индекс оперативности — самый низкий показатель из трех рассмотренных импакт-индикаторов (однако встречаются и исключения, например *Physics of Life Reviews* или *Annals of Family Medicine*, где индекс оперативности, наоборот, выше двух- и пятилетнего импактов).

2.2.3. Средняя цитируемость

Наряду с импакт-фактором, рассчитываемым для журналов, можно ввести аналогичный по сути показатель для оценки средней влиятельности статей отдельного ученого, или лаборатории/факультета, или организации, или целой страны. Необходимо лишь зафиксировать, как и для любого импакт-индикатора, публикационное окно и окно

цитирования исходя из следующей логики. Если изучать только самые свежие статьи, то многие из них еще не наберут достаточное число ссылок, в полной мере отражающее их научный уровень. С другой стороны, если сделать акцент на статьях, вышедших давно, то не удастся оценить текущее состояние исследований автора/организации. Как правило, приемлемым решением является использование пятилетнего публикационного окна и совпадающего с ним окна цитирования. В этом случае считается среднее число ссылок, полученных за последние полные пять лет теми статьями, которые вышли в течение тех же последних пяти лет (в расчете на одну статью).

В ESI информация дается по последним десяти годам, плюс прошедшая часть текущего года (с задержкой в 2 месяца), но существует возможность просмотра отдельно каждого пятилетнего интервала, входящего в десятилетку.

Заметим, что для бурно развивающихся областей, где знание быстро устаревает (таких как нанотехнологии, онкология, прикладная физика), можно выбрать временной интервал меньше пяти лет; для областей, где старение знания происходит особенно долго (математика, зоология, социология), имеет смысл увеличить оба окна. В InCites есть возможность посчитать для организации или автора показатель «*cites per document*» (цитируемость в расчете на одну статью) на любом временном промежутке.

2.2.4. Совокупный и средневзвешенный импакт-фактор

Как следует из определения, импакт-фактор является не характеристикой журнала как целого, а показателем средней влиятельности, среднего уровня *одной статьи*

в журнале. При этом реальный уровень статей, измеренный в полученных ими ссылках, обычно сильно разнится от статьи к статье даже в рамках одного журнала. В журнале с импакт-фактором 2,0 могут встретиться как те статьи, которые в течение двух лет после своего выхода не получили ни одной ссылки, так и те, что получили 20 ссылок (напомним: из определения импакта следует, что в среднем в таком журнале статья получает 4 ссылки за два года, следующих за годом публикации). Импакт-фактор отражает среднюю цитируемость, но не может учесть колебания вокруг среднего.

Именно истинное, реальное число ссылок, полученных статьей, отражает ее подлинную влияние, а импакт-фактор издания, в котором она опубликована, делает это лишь косвенно и недостоверно. Поэтому в п. 2.2.3 мы считали именно реальную цитируемость. Однако у методик, построенных на измерении истинного, «наблюдаемого» числа ссылок, есть серьезный недостаток: для адекватной оценки исследуемых работ необходимо, чтобы после их опубликования прошел значительный промежуток времени. Статьи должны успеть получить то число ссылок, которое отражает их реальный уровень, — в п. 2.2.3 мы рекомендовали с этой целью пятилетний интервал.

Ввиду этого как некоторая оценка именно оперативного состояния научной деятельности организации и — реже — ученого используется показатель, базирующийся на импакт-факторе тех журналов, где выходят их статьи. Такой индикатор может быть посчитан сразу после выхода соответствующей публикации (а в известном смысле даже раньше, в момент принятия статьи в печать). Корректное полное название данного индикатора — «совокупный импакт-фактор журналов, в которых опубликованы статьи

организации/ученого». Рассчитывается он за некоторый промежуток времени, для оперативного оценивания логично брать один (завершившийся) год. Показатель равен сумме импакт-факторов тех журналов, в которых публиковались статьи организации/ученого; если есть несколько статей из одного и того же журнала — соответствующее слагаемое умножается на число статей, вышедших в данном журнале.

Совокупный импакт-фактор характеризует (как и следует из его названия) деятельность организации/ученого «интегрально». Если необходимо ввести оценку в расчете на одну опубликованную статью, используют термин «средневзвешенный импакт-фактор журналов, в которых опубликованы статьи организации/ученого». Последний равен отношению совокупного импакта, который введен выше, к общему числу опубликованных за рассматриваемый промежуток времени статей.

Средневзвешенный импакт-фактор показывает средний уровень статей в тех журналах, в которых публикуется организация/автор, и в какой-то мере позволяет предсказать дальнейшую цитируемость работ организации/автора.

Следует иметь в виду, что совокупный и средневзвешенный импакт-факторы сильно зависят от области знания, в которой работает ученый / функционирует организация. Про нормализацию по областям науки речь пойдет в разделе 2.3. Если же ограничиться рамками одной дисциплины (например, исследовать деятельность отдельного ученого или узкотематической лаборатории), то полученное значение средневзвешенного импакта интересно сравнить с показателем по всем журналам данной дисциплины. Так можно определить, в «сильных» или «слабых» журналах публикуется в сред-

нем ученый/лаборатория. Для иллюстрации можно представить, что организация/автор публикуются все время в одном и том же «среднем» для себя журнале, тогда именно средневзвешенный импакт-фактор отражает место этого «среднего» журнала в иерархии научных периодических изданий соответствующей дисциплины¹⁴.

2.2.5. Роль самоцитирования. Индикаторы, характеризующие самоцитирование

Заметим, что на все показатели, разобранные в настоящем разделе, оказывают влияние и цитирования, сделанные самими «исследуемыми объектами» на себя. В расчет числителя импакт-фактора журнала включаются в том числе ссылки, полученные им из его же статей. При оценке средней цитируемости автора или организации учтены и цитирования автором/организацией своих публикаций и т. д. В своем стандартном определении все рассмотренные в настоящем разделе показатели *включают самоцитирование*.

Нельзя сказать, что учет самоцитирования — это недостаток разобранных метрик или, наоборот, их сильная сторона. Необходимо знать об этой особенности методики и иметь ее в виду. Полезно сравнивать показатели, посчитанные с включенным самоцитированием и с исключенным. Например,

¹⁴Как упоминалось в п. 2.2.1, распределение импакт-факторов сильно асимметрично, с небольшим числом журналов с высокими значениями импакта и множеством изданий с низкими показателями. Это верно в той или иной мере для любой дисциплины. Часто в таких случаях более корректным показателем является медианный импакт-фактор журналов, в которых публикуется ученый/лаборатория, вместо среднего (средневзвешенного) импакта. Соответственно, сравнивать его следует с медианным значением по всей дисциплине.

в JCR для журналов приводится и импакт-фактор (классический, двухлетний), из числителя которого убраны ссылки журнала на самого себя — Impact Factor Without Journal Self Cites.

Для оценки уровня самоцитирования журналов обычно используют два показателя. В числителе обоих индикаторов — число ссылок, полученных журналом из статей, опубликованных в нем самом. Число «самоцитирований». В знаменателе первого показателя число всех ссылок, полученных журналом. Этот показатель называется *коэффициентом самоцитируемости*, он показывает долю во всех ссылках, полученных журналом, ссылок, полученных им из него самого. Знаменатель второго показателя равен числу всех ссылок, сделанных журналом. Он называется *коэффициентом самоцитирования* и показывает долю во всех цитированиях, сделанных журналом, ссылок, ведущих на него самого.

Высокий коэффициент самоцитируемости говорит о том, что журнал почти никто не цитирует, кроме него самого, и это, по заключению, например, Р. Руссо [Rousseau, 2002], свидетельствует о малой заметности журнала. Высокий коэффициент самоцитирования может означать несколько другое: журналу некого цитировать, кроме самого себя. То есть журнал относится к замкнутой, изолированной научной дисциплине. Но если при этом также высок коэффициент самоцитируемости, можно предположить наличие неэтичной установки в политике редакции на цитирование только собственных статей.

Для журналов обычно считается допустимым, еще не свидетельствующим о проблемности издания, коэффициент самоцитируемости в 30–35 %. При этом, как правило, он высок у слабых изданий, получающих мало ссылок, и низок у журналов-лидеров.

Например, для двух российских изданий, данные по которым есть на текущий момент в общественно-научной части JCR-2013, коэффициент самоцитируемости составляет 71% («Социологические исследования») и 75% («Вопросы психологии»). Эти журналы находятся внизу импакт-рейтингов в своих дисциплинах. Если взять журналы-лидеры в соответствующих рубриках JCR-2013, то для них коэффициент самоцитируемости будет 1,9% (American Sociological Review) и 2,9% (Educational Psychologist) соответственно. В частности этот эффект приводит к тому, что исключение самоцитирования при вычислении импакт-факторов журналов (т.е. учет только ссылок, полученных журналом из других изданий), как правило, слабо влияет на рейтинг ведущих журналов с высокими показателями, однако сильно меняет взаимные позиции «на дне» рейтинга, среди малоцитируемых изданий [ср. McVeigh, 2004].

Что касается самоцитирования авторов, для его оценки применяются аналогичные коэффициенты, в первую очередь коэффициент самоцитируемости. Однако в этом случае есть два различных подхода к тому, что именно считать самоцитированием. «Прямое самоцитирование» — это ссылка на публикацию, среди соавторов которой есть исследуемый ученый, появившаяся в статье, среди соавторов которой есть тот же самый ученый. Однако есть еще и самоцитирование соавторов — ссылка на публикацию, среди соавторов которой есть исследуемый ученый, появившаяся в статье, среди соавторов которой есть соавтор(ы) исходной статьи (при этом самого исследуемого ученого в авторах цитирующей публикации может не быть). Поясним подробнее: пусть ученый S написал (сам или в соавторстве) статью P1, которую цитирует статья P2. Это считается:

- прямым самоцитированием, если S — автор P2;
- самоцитированием соавторов, если среди авторов P2 есть хотя бы один из авторов P1.

Соответственно, коэффициент самоцитируемости ученого может считаться или как доля «прямого самоцитирования» во всех ссылках, которые получили его публикации, или как доля «самоцитирования соавторов» во всех ссылках, полученных его работами²⁵. Очевидно, что второй коэффициент будет всегда больше или равен первому. Вообще говоря, исключение при анализе работ ученого всех ссылок, являющихся «самоцитированием соавторов», предлагает весьма ригористичный подход: посчитаем только те цитирования, которые получены из «совсем чужих» работ — множество авторов которых никак не пересекается с множеством авторов цитируемой публикации.

В заключение заметим, что само по себе самоцитирование, конечно, не только не порочная практика, но даже неотъемлемая составляющая научной коммуникации. Журнал не может не ссылаться на свои публикации, это означало бы отсутствие всякой преемственности в его функционировании. Сложно представить себе автора, который бы не ссылался на свои предыдущие работы, — это был бы ученый либо постоянно меняющий область деятельности, либо стыдящийся, отрекающийся от своих прежних публикаций. Но нельзя и не помнить о возможных злоупотреблениях, когда самоцитирование гипертрофируется и искажает библиометрический анализ, если в нем не отслеживать данный аспект.

²⁵Заметим, что в случае с журналами такой вариативности нет по той причине, что одна и та же работа не может быть опубликована сразу в двух журналах, т.е. не бывает «журналов-соавторов».

2.3. Индикаторы относительной влиятельности статей (относительные, нормализованные импакт- индикаторы)

До сих пор мы рассматривали «абсолютные» индикаторы, которые зависят только от показателей самого исследуемого журнала/ученого/организации и не учитывают контекст, в котором те осуществляют свою исследовательскую/публикационную деятельность. Самый серьезный недостаток при этом то, что мы не учитывали, к какой научной области относятся оцениваемые статьи. Сравнение эффективности академической деятельности представителей различных наук невозможно осуществить с помощью абсолютных, «простых» индикаторов. *Центральным фактом, который необходимо иметь в виду при оценке эффективности деятельности ученых или организаций, а также при оценке качества журналов является сильная зависимость абсолютных библиометрических показателей от научной дисциплины, в которой публикуются изучаемый автор/организация или выходит журнал.*

Проиллюстрируем это на журналах. Цитируемость журналов серьезным образом зависит от их дисциплинарной принадлежности, поэтому медиана импакт-факторов сильно варьируется от рубрики к рубрике. Например, для клеточной биологии в JCR-2013 она составляет 3,33, а для математики — 0,58. Это означает, что половина журналов по клеточной биологии имеет импакт выше 3,33, в то время как половина журналов по математике получила импакт ниже 0,58. Более того, в категории «математика» вообще нет ни одного журнала с импакт-фактором выше 3,33 (три таких журнала появятся, если добавить категории «прикладная математика» и «математика, междисциплинарные приложения»).

Почему так происходит? Следует ли говорить о том, что математика имеет какие-то «недостатки» относительно биологических наук, какой-то внутренний изъян? Разумеется, нет. На самом деле можно указать как минимум три причины указанного эффекта:

- Различная практика цитирования, сложившаяся в журналах различных научных областей; можно сказать, варьирующаяся в зависимости от дисциплины «плотность ссылочного поля». Например, известно, что в одной статье по клеточной биологии в среднем 55 ссылок в пристатейном списке литературы, в то время как в математике — лишь 21 (JCR-2013). Поэтому «вероятность» получить ссылку у журнала по математике меньше³⁶. Это индивидуальные особенности коммуникации, укоренившиеся в различных науках и никак не связанные с их «качеством» относительно друг друга.
- Разный средний «возраст» цитируемых источников. В одних дисциплинах основной массив цитирований ведет на свежую литературу (попадающую в двухлетний период, который учитывается при подсчете импакт-фактора), в других областях науки чаще цитируются источники, вышедшие давно (мы упоминали об этом в п. 2.2.2 и поговорим подробнее в разделе 2.7). Чем больше доля ссылок, ведущих на документы старше двух лет, тем меньше они поднимают импакт-факторы журналов соответствующей области науки.
- В некоторых дисциплинах цитируемая литература хорошо представлена в журнальных блоках Web of Science CC, а в некоторых встречается много ссылок на материалы, не охваченные базой данных.

³⁶Разумеется, здесь мы используем приближенный подход, оставляя в стороне междисциплинарность и возможность получения ссылки «извне» дисциплины.

Например, в тех областях, где больше ссылок на книжные, а не журнальные источники; на труды конференций, диссертации, кодексы, интернет-страницы и т. д. [см. Larivière et al., 2006]. Все эти ссылки «пропадают» для импакт-фактора, даже если по своему возрасту цитируемая литература попадает в двухлетнее публикационное окно, — они также не увеличат импакт никакого журнала.

Эти дисциплинарные особенности не учитываются в абсолютном значении импакт-фактора, что делает невозможным его использование в комплексном библиометрическом исследовании. Поэтому для решения двух важных задач: а) сравнение между собой исследовательских единиц, работающих в разных областях науки; б) комплексная оценка деятельности организации, занимающейся исследованиями сразу в нескольких научных областях — вводятся *относительные* библиометрические индикаторы. Цель их — оценить научную деятельность изучаемых объектов *в сравнении* с другими однотипными объектами в той же научной области. Это сравнение «на фоне коллег» или, в обратной формулировке, «относительно конкурентов». Относительные индикаторы потребуются в том случае, если необходимо сравнить достижения химика и математика или получить интегральный показатель вуза, чьи сотрудники публикуют статьи в широком спектре научных дисциплин.

2.3.1. Относительный импакт-фактор, относительная цитируемость

Начнем с относительного показателя для журнала. Самый простой из них — отношение импакт-фактора издания к среднему импакт-фактору дисциплины, к которой журнал относится. При этом среднее по дисциплине обычно берется не как отношение суммы всех

импакт-факторов журналов данной дисциплины к числу таких журналов, а несколько иначе: используется т. н. «агрегированный импакт-фактор» (aggregate impact factor). Дисциплина в целом (т. е. совокупность статей, опубликованных во всех журналах, отнесенных к данной дисциплине) рассматривается как некий единый «метажурнал», для которого вычисляется традиционный импакт-фактор: отношение числа всех ссылок, полученных в году Y статьями данной дисциплины, вышедшими в годах $Y-1$ и $Y-2$, к числу статей данной дисциплины, вышедших в годах $Y-1$ и $Y-2$. Этот показатель для дисциплин как целого также публикуется в JCR. Именно он — «агрегированный» («собранный») импакт-фактор дисциплины — берется в качестве знаменателя, именно на него делится импакт каждого журнала, чтобы получился относительный импакт-фактор.

В целом можно сказать, что если относительный импакт-фактор журнала больше 1, это значит, что в среднем статьи журнала цитируются чаще, чем средняя статья, опубликованная в данной дисциплине, и наоборот. Следует также иметь в виду, что один и тот же журнал может быть приписан сразу к 2–3 дисциплинам и у него, соответственно, может быть посчитано несколько относительных импакт-факторов — относительно агрегированного значения в каждой из дисциплин. В таком случае с помощью описанного метода можно узнать, в какой из научных областей это издание занимает более престижную позицию: в той из дисциплин, у которой ниже агрегированный импакт-фактор, относительный импакт такого журнала, очевидно, будет выше.

Аналогичный показатель можно считать и для авторов. Назовем его для удобства «относительная цитируемость» (хотя это слишком общее название). Относительная

цитируемость ученого равна отношению средней цитируемости его статей (среднее число ссылок на одну статью) к средней цитируемости статей некоей «референтной группы», на фоне которой мы рассматриваем деятельность ученого.

Один из наиболее важных вопросов при библиометрическом анализе — что выбрать в качестве референтной группы? Это могут быть статьи организации, в которой работает ученый, или более узко — статьи его лаборатории/факультета. В первом случае, однако, следует помнить, что если организация публикуется сразу в нескольких областях науки (яркий пример: классический университет), то сопоставление скорее всего будет некорректным: если ученый работает в высокоцитируемой науке, то у него будет незаслуженное преимущество перед коллегами, если в малоцитируемой — он окажется в несправедливо невыгодном положении. Референтная группа должна, по крайней мере, публиковаться в той же научной области, что и оцениваемый автор. Кроме того, как всегда, необходимо четко задавать публикационное окно для оцениваемых статей и окно цитирования — для статей, чьи ссылки учитываются при подсчете индикатора⁴⁷.

⁴⁷При выборе публикационного окна больше одного года проявляется, однако, некоторая слабость предложенного метода. Предположим, что автор публиковал свои статьи не равномерно из года в год, а, например, по нарастающей — все больше и больше, — в то время как во всей его дисциплине выходило примерно одинаковое число статей ежегодно. Тогда автор получает незаслуженную относительную недооценку своих работ, поскольку шансы на получение ссылок у статей, выпущенных позже, меньше, чем у статей, вышедших ранее. То же самое верно для относительного импакт-фактора: если журнал ощутил увеличение число публикуемых статей в тот год, который предшествовал отчетному, он будет проигрывать журналу, не принимавшему таких редакторских решений (ес-

Тот же метод легко обобщается на целую организацию при условии, что ученые в ней публикуются в более-менее одной тематической области. Можно определить среднюю цитируемость статьи данного вуза/института относительно средней цитируемости статьи, публикуемой в той же самой дисциплине всеми российскими учеными или всеми учеными мира. Случай, когда публикационный профиль организации более широк и затрагивает несколько тематических областей, будет рассмотрен в п. 2.3.2.

Отметим, что в качестве референтной группы также могут быть взяты все статьи того же журнала, где публикуется ученый, — в этом случае мы получим оценку, насколько именно его работы выше или ниже по цитируемости, чем средний уровень издания. Преимущество данного метода — в нем практически отсутствует необходимость следить за областью науки: все статьи в журнале чаще всего принадлежат к одной и той же дисциплине (исключение: мультидисциплинарные издания — *Nature*, *Science*, *Proceedings of the National Academy of Sciences*, «Вестник РАН» и т. д.). Если ученый публикуется сразу в нескольких журналах, что более чем естественно, необходим более комплексный подход, который мы также рассмотрим в п. 2.3.2.

Корректную референтную группу можно выбрать несколькими способами, в конце концов этот выбор зависит от задач, которые ставит перед собой специалист, проводящий библиометрическое исследование.

ли предположить, что научный уровень всех статей в обоих журналах одинаков). Эти недостатки преодолеваются при использовании более тщательного подхода, описанного в п. 2.3.2.

2.3.2. Относительная цитируемость разнородного потока публикаций («normalized citation impact», «crown indicator»)

Представим себе теперь ситуацию, максимально приближенную к реальной: необходимо оценить научный уровень статей, опубликованных организацией широкого научного профиля, например университетом.

Сначала, конечно, необходимо задать публикационное окно и окно цитирования, чтобы определить, какие именно публикации и на каком временном интервале мы оцениваем — так, как это делалось в п. 2.3.1. Здесь особенных сложностей не возникает. Однако выделенный публикационный поток будет крайне неоднородным: он будет содержать публикации из разных журналов и разных тематических областей, кроме того это будут публикации разных типов (исследовательские статьи, научные обзоры, письма и т. д.), это приводит к дополнительным сложностям.

В данном случае необходимо исходить из двух основных положений, которые уже фигурировали в п. 2.3.1:

— определять «научный уровень» можно только на некотором фоне, в некотором контексте, *относительно* чего-то; необходимо задать *референтную группу*;

— сравнивать научный уровень публикаций относительно друг друга можно, только если они *однотипны*, т. е. *все их характеристики, не имеющие прямого отношения к научному уровню, одинаковы*.

Если мы анализируем позиции факультета внутри вуза, то референтной группой будет весь вуз. Если исследуется университет как целое, то в качестве референтной группы логично выбрать все российские вузовские публикации (заметим: включая публикации и самого анализируемого универ-

ситета) — если мы хотим оценить эффективность научной деятельности университета на фоне других вузов России. Можно также взять вообще все отечественные работы — так мы узнаем, как выглядит университет на фоне всей российской науки.

Однако для унификации процедуры оценки мы в дальнейшем будем считать референтной группой весь мир, т. е. все мировые публикации. Таким образом, будет оцениваться соответствие исследуемого научного коллектива общемировому уровню. Именно так построена работа в продукте InCites: можно найти показатели автора, университета или целой страны относительно среднемирового уровня⁵⁸.

На цитируемость публикации влияет ряд факторов, не имеющих непосредственного отношения к ее научному уровню:

— область науки (как мы видели, средняя цитируемость сильно зависит от научной дисциплины);

— год издания (чем раньше издана статья, тем больше ссылок она могла получить к моменту измерения);

— тип публикации (как уже говорилось в п. 2.1.1, в среднем, например, научные обзоры цитируются чаще, чем оригинальные исследовательские статьи⁶⁹).

⁵⁸Заметим, что этот подход не замещает собой оценку более компактных групп относительно друг друга. Например, если посчитаны показатели одной организации и целой страны относительно всего мира, то индикатор, показывающий эффективность деятельности этой организации на фоне этой страны, не равен отношению двух найденных показателей друг к другу.

⁶⁹Этот факт далек от чистой теории. Например, документы типа Article, опубликованные в журнале Tetrahedron в 2010 г., получили за 2010–2013 гг. (согласно Web of Science CC) в среднем по 9,0 ссылок. Тот же показатель для публикаций типа Review равен 35,2 — разница почти в 4 раза! Учет типа документа не менее важен, чем коррекция на область науки, к которой он относится.

Для учета всех этих факторов и корректного определения относительной цитируемости всего публикационного потока организации/автора вводится показатель «ожидаемое число ссылок» (expected citation rate, или baseline): для каждой отдельной статьи исследуемых организации/автора это среднее число ссылок, которое получили все мировые публикации, имеющие *ту же дисциплинарную принадлежность, тот же год выхода и тот же тип документа*, что и рассматриваемая статья. Если рассматриваемая статья организации/автора в действительности получила больше ссылок, чем ожидаемое для нее их число, то она превосходит средний мировой уровень «такой же, как она» публикации (т. е. имеющей те же характеристики дисциплина / год выхода / тип документа), если меньше — уступает ему.

Относительная цитируемость одной публикации, таким образом, равна отношению реального («наблюдаемого») числа полученных ею ссылок к их ожидаемому числу. Для того чтобы посчитать относительную цитируемость совокупности всех публикаций организации/автора, берется среднее арифметическое значений относительной цитируемости для каждой отдельной публикации, написанной организацией/автором. Следуя терминологии InCites2, обозначим этот показатель «Normalized Citation Impact» (NCI)⁷¹⁰.

⁷¹⁰В западной литературе данный или аналогичный показатель еще называют «mean normalized citation score», «normalized mean citation rate», «normalized impact» и даже «crown indicator» (подчеркивая его важность). Иногда используется альтернативный способ усреднения по всему массиву ссылок, полученных публикациями организации/автора. Сначала считается ожидаемое число ссылок сразу для всех статей (суммированием по каждой), потом на эту сумму делится суммарное наблюдаемое число ссылок, полу-

В методику расчета показателя NCI заложена чрезвычайно грамотная схема, позволяющая максимально корректно оценить относительную цитируемость любого набора статей. Однако и этот индикатор — как абсолютно любой библиометрический показатель, взятый сам по себе, — имеет свои ограничения. Поскольку он так или иначе считает средние величины, на него могут оказывать очень большое влияние отдельные высокоцитируемые статьи организации/автора. Даже для целых стран, как показали Aksnes & Sivertsen [Aksnes, Sivertsen, 2004], отдельные работы, получившие очень много ссылок, могут значительно повлиять на средние страновые показатели. С не меньшей осторожностью следует применять NCI на коротких временных промежутках, например для статей 1–2-летней давности. Ввиду того что весь массив мировых публикаций получил еще недостаточно много ссылок к этому моменту, показатели ожидаемой цитируемости могут быть мало-предсказуемы и не давать надежного, достоверного базиса нормализации.

Наконец, как уже говорилось ранее в сноске, NCI, полученный относительно мирового уровня, не всегда дает возможность сравнивать исследуемые единицы между собой. Как рельефный пример — если у организации NCI = 0,5, а у всей России NCI = 0,67 (последняя цифра — реальное значение из InCites2 для статей 2010–2014 гг. выхода, с тем же окном цитирования), это не обязательно означает, что организация работает слабо относительно всей российской науки. Возможно, это химический институт,

ценное публикациями организации. Этот способ используется в InCites1 (под названием «Category Actual/Expected Citations»), и именно так определяется показатель «цитируемость, нормализованная по областям науки» [Писляков, 2014].

публикующий статьи лишь в два раза слабее среднемирового уровня, в то время как позиция России в химии ощутимо ниже (0,40 от среднемирового уровня цитируемости). При этом значение NCI = 0,67 для всей страны получается из-за более качественных публикаций в физике⁸¹¹.

2.3.3. Нормализация по журналам и комбинированный анализ относительной цитируемости

Подход, аналогичный представленному в п. 2.3.2, также возможен, если в качестве референтной группы брать публикации не в рамках одной и той же научной области, а в одних и тех же журналах. Тогда «ожидаемое число ссылок» для каждой публикации исследуемых организации/автора будет равно среднему числу цитирований, полученных публикациями данного журнала, которые вышли в том же году и относятся к тому же типу документа. Наблюдаемое число ссылок для статьи, естественно, остается без изменений. Необходимо взять его отношение к ожидаемому показателю и усреднить результат по всем статьям организации/автора. Полученный индикатор, вновь используя термины InCites2, назовем Journal Normalized Citation Impact (JNCI)⁹¹².

⁸¹¹ Данный пример иллюстративен, но охватывает лишь часть указанной проблемы: так, найдя значение NCI для России с ограничением по химическим наукам, все равно некорректно сравнивать его с NCI отдельной организации, специализирующейся на химии: необходим комплексный подход с учетом и годов выхода статей, и типов документа. Такой показатель предусмотрен в плане развития платформы InCites.

⁹¹² Здесь вновь возможна иная нормализация, с подсчетом сразу всех ожидаемых ссылок, суммой по всем статьям, а затем деление на полученный результат суммы всех полученных в реальности ссылок. Это соответствует метрике «Journal Actual/Expected Citations» в интерфейсе InCites1 и показателю «цитируемость, нормализованная по журналам» [Писляков, 2014].

Этот индикатор характеризует то, как «выделялись» статьи данной организации на фоне всех статей тех журналов, где публиковались ее сотрудники. Были публикации ее ученых в среднем выше среднего уровня публикующих их журналов – или ниже него. Тот же самый смысл имеет JNCI, если мы исследуем публикации отдельного ученого или, наоборот, целой страны. Показатель, посчитанный для целых стран, называют также по-другому – «индекс Матфея». Его подсчет дает повод рассуждать о том, «дискриминируются» ли в плане цитируемости статьи ученых из разных стран в рамках одного и того же журнала [Писляков, Дьяченко, 2009; Pisyakov, Dyachenko, 2010].

Нормализация по журналам в известном смысле легче и точнее: как мы уже говорили, здесь почти не встает вопрос об определении тематической области, к которой относится публикация. В рамках одного журнала, за некоторыми исключениями (Science, Nature и т. д.), все статьи принадлежат к одной и той же дисциплине. Или, по крайней мере, эти отличия будут в среднем меньше, чем погрешность при любом другом автоматическом определении тематики публикации. Более того, если сами дисциплины приписываются статьям на основе тематической рубрикации журналов, в которых они опубликованы (как обычно делается в библиометрических базах данных), то получается, что нормализация по журналам дает более точный, детальный взгляд на сопоставление публикаций, чем нормализация по областям науки.

Здесь, однако, есть ключевой момент. Предположим, что некоторый университет получает высокий показатель нормализованной по журналам цитируемости, JNCI. Но при этом, если мы посчитаем его цитируемость, нормализованную по областям

науки, NCI, может оказаться, что она меньше единицы. Это происходит в том случае, если авторы данного вуза выбирают для публикации своих статей слабые журналы. В рамках этих слабых журналов они могут превзойти уровень остальных статей, так что цитируемость их публикаций превысит средние показатели этих непрестижных изданий (и JNCI будет больше 1). Но из-за того что сами журналы мало цитируются на фоне других изданий в той же области науки, цитируемость статей авторов вуза не достигнет среднего уровня по дисциплине.

Конечно, возможна и обратная ситуация, когда авторы организации подают свои статьи в ведущие журналы, успешно проходят процедуру рецензирования и отбора, одна-

ко, попав туда, их публикации не достигают среднего уровня цитируемости этих изданий. В таком случае мы можем получить высокую цитируемость относительно областей науки и низкую — при нормализации по журналам.

Так или иначе очевидно, что для всестороннего исследования научной деятельности организации или автора (а также научных подразделений, целых регионов, стран и т. д.) следует использовать оба показателя — и NCI, и JNCI. Суммируем выводы, которые можно сделать при их совмещенном анализе, в таблице 3. В качестве примера будем рассматривать публикации организации, хотя аналогичные результаты можно получить, исследуя другие научные единицы.

Таблица 3

Анализ нормализованных показателей цитируемости для организации

	JNCI > 1	JNCI < 1
NCI > 1	В среднем публикации сотрудников таких организаций цитируются выше среднего уровня цитируемости как журналов, в которых они выходят, так и областей науки, к которым они относятся	В среднем публикации сотрудников таких организаций цитируются ниже среднего уровня цитируемости журналов, в которых они выходят, но при этом выше уровня областей науки, к которым они относятся. Это может быть связано с выбором сотрудниками сильных журналов для своих публикаций ¹³
NCI < 1	В среднем публикации сотрудников таких организаций цитируются выше среднего уровня цитируемости журналов, в которых они выходят, однако ниже уровня областей науки, к которым они относятся. Это может быть связано с выбором сотрудниками слабых журналов для своих публикаций	В среднем публикации сотрудников таких организаций цитируются ниже среднего уровня цитируемости как журналов, в которых они выходят, так и областей науки, к которым они относятся

¹³Осторожная формулировка «может быть связано» вызвана тем, что при нормализации, используемой в InCites2, сложно определить понятие «среднего уровня цитируемости совокупности журналов, в которых публикуются ученые организации, относительно среднего уровня цитируемости областей науки, к которым относятся эти журналы». Точнее, характеристика «силы» (научного уровня) совокупности журналов, в которых публикуются сотрудники, не определяется однозначным образом значениями NCI и JNCI для организации. При использовании другого способа нормализации [Писляков, 2014] понятие «среднего уровня совокупности журналов относительно дисциплин» может быть формализовано более четко.

В дополнение к характеристикам, представленным в таблице 3, отметим, что показатель $JNCI < 1$ можно трактовать и в положительном смысле для организации: ее сотрудники сумели опубликоваться в журналах, уровень которых выше, чем их конкретные статьи. В то же время $JNCI > 1$ может означать, что ученые регулярно «недооценивают» свои работы и, быть может, им стоит задуматься о подаче рукописей в журналы более высокого уровня (последнему наблюдению я обязан О. В. Москалевой (частная беседа).

Исследование NCI и $JNCI$ используется для анализа цитируемости статей и публикационной стратегии организаций как в фиксированный момент времени, так и в динамике. В последнем случае изменение показателей организаций отражает направление и темп эволюции научной деятельности ученых соответствующих вузов/институтов/лабораторий и т. д.

2.3.4. Ранговые индикаторы

Рассмотрим еще один подход, используемый при проведении кросс-дисциплинарных сопоставлений журналов. Можно принципиально отказаться от исследования абсолютных значений библиометрических показателей, а сравнивать лишь места, позиции в рейтинге, которые занимают издания в своих дисциплинах. После сортировки списков журналов в двух различных дисциплинах в порядке убывания какого-либо выбранного библиометрического индикатора (например, импакт-фактора, пятилетнего импакт-фактора; или индекса влияния, который будет введен в разделе 2.5) проводится анализ только занятых в этой иерархии изданиями порядковых мест.

Суть рангового метода — разбиение полученного упорядоченного списка-рейтинга

на n равных частей и определение, в какую из этих частей попадают журналы. Например, если n принимается равным 4, тогда говорят о квартилях — журналы первого квартиля (попадающие в верхнюю четверть списка по рассматриваемому параметру), второго квартиля, третьего квартиля, четвертого квартиля. В системе ранговых индикаторов считается, что журналы, попавшие в первый квартиль, выше журналов, попавших (не только в этой, но и в любой другой дисциплине) во второй квартиль и т. д. При этом журналы, попавшие в один и тот же квартиль в различных дисциплинах, считаются в рамках этого метода одинаковыми по научному уровню.

Как выбирать показатель n — непростой вопрос, и ответ на него зависит от конкретной задачи. Фактически, n задает «точность», «разрешающую способность» рангового индикатора. Журналы, оказавшиеся равными при выборе $n = 4$ (попавшие в один квартиль), могут получить разный уровень при $n = 5$ («квintиль») и т. д. И хотя чем больше n , тем выше «разрешение» рангового метода, у более грубого подхода, у деления на те или иные более крупные «процентили» (части), есть свое преимущество: некоторые группы журналов попадают в один и тот же процентиль, что интуитивно соответствует представлению о том, что существуют «примерно одинаковые» издания.

Существует предельный случай «уточнения» рангового метода, когда весь ряд занятых журналами мест от 1 до N (N — число журналов в рассматриваемой дисциплине) проецируется на отрезок $[0; 1]$, и каждый журнал оказывается на своей точке этого отрезка. Несложно вычислить, что при этом показатель «относительной позиции» журнала, занимающего n -е место в своем дисциплинарном рейтинге, равен $(N - n) / (N - 1)$.

При кросс-дисциплинарном сравнении «выигрывает» тот журнал, у которого указанная «относительная позиция» выше. Это показано на рис. 10, где сравниваются журналы двух различных дисциплин. В одной из них (дисциплина А) 5 изданий, соответственно они занимают места с 1-го по 5-е в рейтинге по своей дисциплине, во второй рубрике (дисциплина В) 7 изданий с местами с 1-го по 7-е. После проецирования на отрезок $[0; 1]$ итоговое кросс-дисциплинарное ранжирование выглядит так (в угловые скобки заключены издания, занявшие одинаковые места): $\langle A_1, B_1 \rangle - B_2 - A_2 - B_3 - \langle A_3, B_4 \rangle - B_5 - A_4 - B_6 - \langle A_5, B_7 \rangle$.

Аналогичным образом та или иная модификация рангового метода может быть применена при сравнении как авторов, так и организаций между собой. При этом важно корректно и полно составить ту группу, объединяющую работающих в одной научной области, внутри которой будет определено взаимное место авторов/организаций (аналог научной дисциплины для журналов). В случае с авторами или узкоспециализированными организациями это сделать проще, в случае с организациями широкого профиля — сложнее, в этом случае может оказаться, что ранговые подходы неприменимы.

В заключение отметим, что одно из преимуществ ранговых методов — защита от сильной асимметричности распределений цитирования, когда небольшое число журналов (авторов, организаций) получают значительное число ссылок и создают такие средние показатели по своей группе, что все остальные участники не могут их достичь (ср. сноску п. 2.2.4, где ранговый подход, а именно использование медианы, предлагался для анализа импакт-факторов). В этом случае отказ от учета абсолют-

ных значений индикаторов и исследование лишь мест в рейтингах может оказаться более эффективным.

2.4. Индикаторы «экстремальной» цитируемости

Еще один подход, коренным образом отличающийся от разобранных выше, заключается в том, чтобы не рассчитывать среднюю цитируемость всего массива статей, а заниматься «крайностями», наблюдаемыми

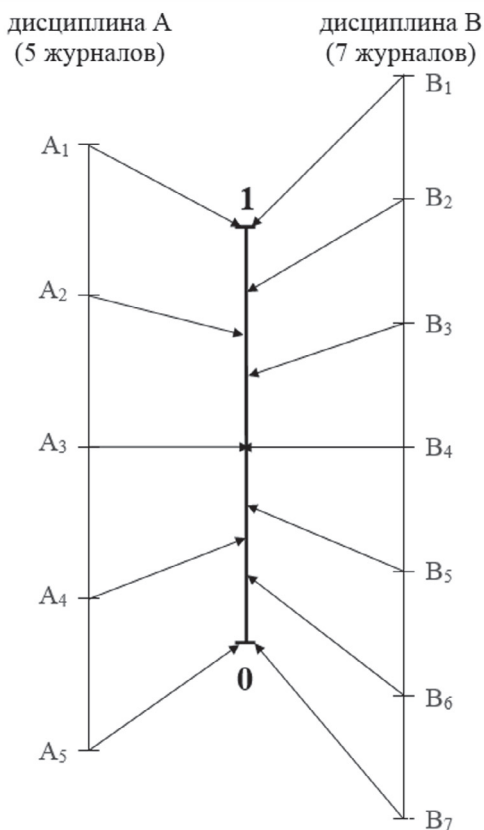


Рис. 10. Сравнение рангов журналов проецированием на отрезок $[0; 1]$.

A1... A5 — журналы дисциплины А.
B1... B7 — журналы дисциплины В

в нем, — теми статьями, которые получили аномально малое или аномально высокое число ссылок. Поэтому назовем показатели, основанные на такой методике, индикаторами «экстремальной» цитируемости, имея в виду «экстремум», предел возможной цитируемости как со стороны максимума, так и со стороны минимума. Со стороны минимума очевидным образом фигурируют публикации, не получившие ни одной ссылки. Со стороны максимума — т. н. «высокоцитируемые» статьи (*highly cited papers*), определение которых дается в п. 2.4.1.

Отличительной чертой «экстремально-го» подхода является то, что определяется лишь доля статей, попавших в две «предельные» категории, а число реально полученных ими ссылок не учитывается (впрочем, в случае минимального цитирования это число известно — 0). Важен вопрос, сколько в массиве высокоцитируемых и нецитируемых статей, каков их процент. Оба метода могут применяться по отдельности, они независимы друг от друга, необязательно использовать их в паре.

2.4.1. Высокоцитируемые статьи

В библиометрии использовались различные способы определения высокоцитируемых статей. Это могли быть и статьи, получившие определенное абсолютное число ссылок (данный метод, естественно, не может использоваться при анализе мультидисциплинарного множества публикаций), и статьи, занявшие высокие места по цитируемости в рамках определенной группы (например, «лучшие (по числу полученных ссылок) 10 публикаций университета»). Однако сейчас наиболее распространено определение, использующее «процентили» — верхний срез в n % наиболее цитируемых статей. В зависимости от ис-

следовательских задач показатель n может варьироваться, мы разберем подход, используемый в базе данных ESI.

В рамках ESI высокоцитируемой считается публикация, которая попадает в 1 % наиболее цитируемых работ в мире среди тех, которые вышли в том же, что и она, году и в той же области науки. Для каждого года выхода и каждой области науки в ESI высчитываются пороговые значения числа цитирований, которое должна получить к настоящему моменту статья, чтобы попасть в 1 % высокоцитируемых. Как и ранее в п. 2.3.2, год публикации фиксируется для того, чтобы убрать преимущество статей, вышедших давно, перед более свежими публикациями: у последних физически меньше времени набрать большое число цитирований. Ограничение по научной дисциплине, как мы знаем, «уравнивает» статьи, вышедшие в областях науки с высоким и низким средним уровнем цитируемости. Например, по данным на сентябрь 2014 г. высокоцитируемой статьей 2004 г. выхода по дисциплине «микробиология» была лишь та, которая получила за прошедшие 10 лет не менее 214 цитирований, в то время как статье по математике достаточно было получить 68 ссылок. Аналогичные пороги для публикаций 2009 г. выхода по этим дисциплинам составляли 116 и 34 ссылки соответственно¹⁰¹⁴.

Следует иметь в виду, что «высокоцитируемость» — динамическая категория, и статья, однажды ставшая высокоцитируемой, может в дальнейшем потерять этот статус. Это произойдет, если активность ее цитирования в дальнейшем сократится и ее

¹⁰¹⁴Заметим, что, в отличие от методов, разобранных в п. 2.3.2, нормализация не затрагивает тип документа. При этом в ESI обрабатываются только документы типа Article и Review.

обойдут другие публикации того же года выхода из той же области науки. Возросший с течением времени порог, необходимый для попадания статьи в высокоцитируемые, может превзойти число реальных ссылок, полученных данной статьей. Очевидно, что возможен и обратный процесс — статья, не являвшаяся высокоцитируемой, может начать получать много ссылок и попасть в престижную категорию. Поэтому анализ высокоцитируемых статей надо проводить в некоторый фиксированный момент, на множестве публикаций, являющихся в настоящее время высокоцитируемыми. Далее исследование можно периодически повторять, отслеживая динамику.

При анализе высокоцитируемых статей исследуется прежде всего абсолютный показатель — сколько таких работ опубликовал ученый, или лаборатория, или институт, или даже целая страна. Это важный результат, след участия не просто «в науке», но «в науке высочайшего уровня». Такие статьи можно называть научными шедеврами [Писляков, 2011].

Но еще более важным является относительный индикатор, показывающий, какова доля статей, написанных той или иной исследовательской единицей, которые становятся высокоцитируемыми. Может быть, каждая 20-я статья, написанная организацией/ученым, становится высокоцитируемой? Или, может быть, лишь каждая 200-я попадает в эту категорию? При этом здесь существует четкий ориентир, контрольный уровень, относительно которого можно сравнивать эффективность деятельности научного коллектива на мировом фоне. Поскольку, согласно определению, в мире каждая сотая статья является высокоцитируемой, то, если 1 % публикаций научного коллектива (каждая сотая статья) достига-

ет категории высокоцитируемости, это говорит о соответствии деятельности ученых мировому уровню¹¹⁵.

При реализации на практике следует помнить, что высокоцитируемыми становятся лишь 1 % статей, следовательно, анализ высокоцитируемости можно проводить только на больших публикационных массивах, охватывающих тысячи исходных документов. В противном случае выводы исследования будут слишком зависимы от случайных выбросов.

2.4.2. Коэффициент нецитируемости

Коэффициент нецитируемости, как следует из его названия, это доля научных работ, которые ни разу не были процитированы. Результаты научной деятельности научного коллектива тем эффективнее, чем *меньше* значение данного коэффициента. С некоторой степенью огрубления можно говорить, что чем ниже коэффициент нецитируемости, тем меньше «бесполезных», «бесплодных» публикаций выходит из-под пера сотрудников рассматриваемого коллектива¹²¹⁶.

Теперь предположим, что некоторая статья не получила на текущий момент

¹⁵Строго говоря, ситуация несколько сложнее. Высокоцитируемые статьи часто пишутся в больших соавторских коллективах, и для корректного соотношения с мировым уровнем следует использовать дробный счет (fractional counting, см. п. 2.1.1). Точнее будет сказать, что если организация пишет статьи, становящиеся *в менее 1 % случаев* высокоцитируемыми, то она заведомо проигрывает на общемировом фоне. Это порог снизу, а вот обратное часто неверно.

¹⁶Зеркальный взгляд на это явление реализован в InCites2, где считается показатель %Documents Cited, доля процитированных хотя бы один раз документов. Коэффициент нецитируемости, выраженный в процентах, равен 100 % минус %Documents Cited. Но следует иметь в виду, что в InCites2 используется не нормализованный по времени показатель (т. е. не такой, как вводится далее в данном разделе).

ни одной ссылки. Очевидно, что нельзя предугадать, будет ли она процитирована в неограниченной перспективе. Поэтому для грамотного анализа коэффициента нецитируемости выделяется некоторый временной интервал, величину которого фиксируют для публикаций любого года выхода, с тем чтобы более старые статьи не имели преимущества перед вышедшими недавно. Например, если в качестве такого интервала выбран пятилетний отрезок, то производится поиск статей, опубликованных в 2004 г. и не процитированных до 2008 г. включительно, опубликованных в 2005 г. и не процитированных до 2009 г. включительно и т. д. Это позволяет провести корректный анализ динамики коэффициента нецитируемости и определить, например, что доля ни разу не процитированных публикаций организации/автора с 2004 по 2009 г. увеличилась или уменьшилась. Для сравнения различных организаций или авторов между собой необходимо оставаться в рамках одного дисциплинарного направления, поскольку среднемировой коэффициент нецитируемости, как и остальные ненормализованные показатели цитируемости, сильно зависит от рассматриваемой области науки.

В отличие от анализа высокоцитируемости, исследование другого показателя «экстремальной» цитируемости не требует очень больших массивов данных. Например, как было показано [Писляков, 2010], пятилетний коэффициент нецитируемости всех российских статей, вышедших в 2004 г. (по SCIE и SSCI), составил около 40 %. Это означает, что исследуемому массиву публикаций достаточно содержать сотни или даже десятки статей.

2.5. «Взвешенные» индикаторы. Собственный фактор (Eigenfactor) и индекс влияния статьи (Article Influence)

Все изученные ранее индикаторы имели одну общую черту: ссылки, которые получали журнал/автор/организация, были для них «одинаковы», показатели «не замечали», откуда именно получено то или иное цитирование. Замена одного цитирующего журнала на другой, который дает столько же ссылок, не привело бы к изменению ни одного индикатора.

Следующие два индикатора называются «взвешенными», поскольку они учитывают полученные журналом цитирования с различным весом — в зависимости от того, насколько «влиятелен» тот источник, из которого получено цитирование. Эта влиятельность, в свою очередь, зависит от цитируемости самого источника ссылки. Таким образом, ссылка из высокоцитируемого журнала, например Nature, «ценится» выше, чем ссылка из малоизвестного регионального издания. Аналогичная идея лежит в основе расчета ранга «авторитетности» веб-страниц некоторыми поисковыми машинами, например Google (т. н. «PageRank»).

Взвешенный журнальный индикатор, получивший название «собственный фактор» (Eigenfactor), был предложен в 2007 г. специалистами из лаборатории Карла Бергстрёма (Университет Джорджа Вашингтона). Разработчики данного показателя иллюстрировали его с помощью процесса случайного перехода по ссылкам: некий читатель переходит от журнала по случайно выбранной в нем библиографической ссылке к другому журналу. Для соединения не связанных между собой ссылками изданий дополнительно вводится некоторая ненулевая вероятность того, что читатель выберет

следующий журнал случайным образом, а не путем следования по ссылке. Итоговый показатель «престижа» издания будет пропорционален числу заходов читателя в соответствующий журнал.

Данный процесс реализует описанную выше идею о разной ценности ссылок из журналов с разной цитируемостью. Ведь для того чтобы воображаемый читатель часто заходил в некий журнал, необходимо не только чтобы «путей» (ссылок) к этому журналу было много, но и чтобы приходили эти ссылки из часто посещаемых читателем журналов.

Еще один механизм, которым иллюстрируют взвешенные показатели, — это «задача престижа» журналом в результате итеративной процедуры. Сначала каждому журналу назначается одинаковый начальный «престиж», равный единице. На первом этапе итерации каждый журнал «делит» свой единичный престиж между всеми ссылками, которые он «посылает» другим журналам. Чем меньше в журнале сделано ссылок, тем больше «вес» каждой из них¹⁷. Таким образом рассчитывается «престиж», полученный каждым журналом из других журналов. Это значение престижа, полученного в результате первой итерации, принимается в качестве нового значения престижа журнала, которое на второй итерации вновь «делится» между всеми ссылками, исходящими из журнала. Далее процесс повторяется до достижения квазистабильного состояния, когда очередной шаг итерации уже практически не меняет значения «престижа», полученного каждым из журналов.

¹⁷Точно так же в случае случайного перехода по ссылкам: чем меньше ссылок в журнале, тем больше вероятность, что воображаемый читатель воспользуется какой-то конкретной из них, а значит, выше ценность каждой ссылки.

(Точный алгоритм устроен таким образом, что данное состояние будет обязательно достигнуто.)

Согласно строгому математическому определению, для вычисления значения собственного фактора решается матричное уравнение и находится собственный вектор матрицы вероятностей переходов (отсюда название «собственный фактор»). Собственный фактор — показатель журнала как целого. Если его нормировать на число статей в издании, получится индикатор, характеризующий среднюю статью в журнале и называемый «индексом влияния статьи» (Article Influence). Оба показателя публикуются с 2007 г. в JCR.

При вычислении собственного фактора журнала и индекса влияния статьи учитываются ссылки текущего года на публикации, вышедшие за пять предыдущих лет, — публикационное окно и окно цитирования совпадают с установленными для пятилетнего импакт-фактора. Отличительной особенностью этих показателей является то, что ссылки журнала на свои же статьи не учитываются, — самоцитирование при расчете исключается.

Использование взвешенных индикаторов в совокупности с традиционным импакт-фактором (двух- или пятилетним) позволяет взглянуть на цитируемость журналов более объемно. В частности, при их сопоставлении, как правило, рельефно выделяются журналы, которые получили высокие импакт-показатели благодаря самоцитированию или ссылкам (зачастую взаимным) из узкого, замкнутого круга изданий. При высоком значении импакт-фактора (импакт-факторов) эти журналы будут иметь низкие значения взвешенных индикаторов: такого рода «островки взаимного цитирования» не дадут высокие

значения собственному фактору журнала / индексу влияния статьи. В модели «случайного блуждания» читатель редко будет оказываться на этих островках, если они не связаны с остальными журналами полученными из них ссылками (а кроме того, из рассмотренных нами взвешенных индикаторов вообще исключено самоцитирование). Например, один из журналов, уличенный в практике «взаимного договорного цитирования», в JCR-2012 занимал 14-е место по импакт-фактору в дисциплине «экономика» — и при этом 197-е(!) место по индексу влияния статьи.

Наконец, отметим, что индекс влияния в какой-то степени сглаживает разницу уровней активности цитирования в различных научных дисциплинах (главная проблема при использовании абсолютных значений импакта), поскольку «престиж» журнала распределяется поровну между всеми исходящими из него ссылками. В дисциплинах, где среднее число ссылок в одной статье велико, вес одной ссылки будет меньше, чем в тех областях науки, где ссылок в расчете на статью меньше и «ссылочное поле» менее плотное. Кроме того, пятилетнее публикационное окно также способствует более равновесному учету цитирований в различных научных дисциплинах.

2.6. Индекс Хирша (h -index) и «хирш-подобные» показатели

Индекс Хирша был предложен физиком из Калифорнийского университета в Сан-Диего Х. Хиршем [Hirsch, 2005]. Этот показатель кардинальным образом отличается от всех рассмотренных ранее, поскольку он пытается дать комплексную оценку одновременно числу публикаций ученого и их цитируемости (т. е. «количеству и качеству»). Кроме того, методика подсчета

индекса Хирша намеренно уходит от определения средних величин, средней цитируемости статей (в этом он схож с методиками «экстремальной» цитируемости, разобранными в разделе 2.4). Это имеет свой смысл, поскольку усреднение зачастую не дает полной картины об исследуемом множестве статей и не позволяет производить корректное сравнение эффективности научной деятельности авторов и организаций. При усреднении сильные искажения могут быть вызваны отдельными публикациями, которые получают очень много ссылок, являясь фактически «выбросами», выделенными случаями.

Индекс Хирша имел невероятный успех в научном сообществе, вдохновив специалистов в библиометрии на создание многих десятков своих «модификаций» и «усовершенствований». Мы рассмотрим лишь два примера этих «хирш-подобных» метрик, однако начнем с рассмотрения классического индекса.

2.6.1. Индекс Хирша

Согласно определению, индекс Хирша массива публикаций ученого равен h , если есть h статей из этого массива, каждая из которых получила не менее h цитирований, а каждая из остальных статей при этом получила не более h цитирований. Эквивалентное определение: индекс Хирша массива публикаций ученого равен h , если есть h статей из этого массива, каждая из которых получила не менее h цитирований, но нет $h + 1$ статей, каждая из которых получила бы не менее $h + 1$ цитирований.

Например, если у ученого есть 5 статей, которые получили 10, 8, 2, 2 и 0 ссылок соответственно, то его индекс Хирша равен 2, поскольку есть 2 статьи, каждая из которых цитировалась не менее 2 раз (таких статей

даже 4), но уже нет 3 статей, каждая из которых цитировалась бы не менее 3 раз¹⁴¹⁸.

Графически это проиллюстрировано на рис. 11. Каждая точка на графике соответствует одной публикации ученого. Публикации упорядочены по цитируемости: от тех, которые получили больше всего ссылок, к наименее цитируемым статьям. По оси абсцисс на рис. 11 отложен порядковый номер публикации в этом упорядоченном списке. По оси ординат — сколько ссылок получила статья с соответствующим номером. Ввиду упорядочения по цитируемости, у точек постепенно уменьшается ордината при увеличении абсциссы. Если мысленно соединить точки для получения графика, то h -индекс будет находиться в районе пересечения этого графика с показанной на рисунке биссектрисой $y = x$. Точнее, индекс Хирша будет соответствовать максимальному номеру статьи, находящейся *не ниже* прямой $y = x$ (выше или на этой прямой). Как видно из геометрии графика, у рассматриваемого ученого есть h статей, каждая из которых цитируется не менее h раз, но при этом у него нет $h + 1$ статей, каждая из которых цитируется не менее $h + 1$ раз: статья с номером $h + 1$ находится ниже прямой $y = x$, а значит получила менее $h + 1$ ссылок.

Основное свойство индекса Хирша: увеличение публикационной активности без достаточной цитируемости публикуемых работ не приведет к его росту. И наоборот: получение большого числа цитирований одной-двумя статьями также не повысит индекс серьезным образом. Если автор на-

писал 100 статей, но при этом каждая из них получила 5 ссылок, индекс Хирша ученого будет равен всего лишь 5. С другой стороны, если автор написал лишь 5 статей, каждая из которых получила по 100 ссылок, его индекс все равно будет равен 5. Для достижения высокого значения индекса Хирша необходимо, чтобы автор писал *много таких работ, каждая из которых получала бы много цитирований*. Именно в этом смысле h -индекс пытается отразить сбалансированную оценку сразу и публикационной активности ученого, и цитируемости его работ.

Введенный изначально для авторов, индекс Хирша может быть распространен на организации, страны и вообще любой набор статей, для каждой из которых известно число полученных ссылок.

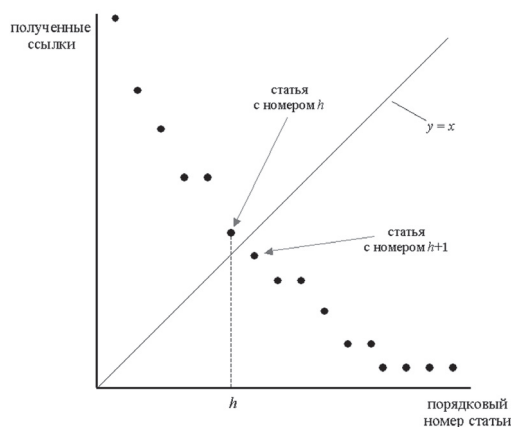


Рис. 11. Индекс Хирша (h -индекс)

Для ученых часто считают «пожизненный» индекс Хирша за всю научную карьеру (как было сделано Х. Хиршем в исходной статье), однако это приводит к очевидному неравенству пожилых и молодых специалистов. Дело в том, что такой показатель не может уменьшаться с течением времени,

¹⁴¹⁸В строгом математическом виде определение индекса Хирша формулируется так: пусть у ученого есть N статей, упорядочим их от наиболее цитируемых к наименее цитируемым. Обозначим c_i число ссылок, полученных i -й статьей, $c_i \geq c_{i+1}$ для всех $i = 1 \dots N - 1$. Тогда индекс Хирша $h = \max\{i: c_i \geq i\}$.

только расти или оставаться неизменным¹⁵¹⁹, — и ученый может уже много лет находиться на пенсии и не писать научных трудов, однако h -индекс его будет не меньше, чем он был на пике карьеры.

Поэтому для получения более осмысленного показателя следует, как и для любой библиометрической величины, задать публикационное окно и окно цитирования. Например, можно рассматривать все статьи, вышедшие у ученого за пятилетний промежуток, и учитывать ссылки, полученные этими статьями в течение того же пятилетнего интервала.

Следует упомянуть еще о некоторых ограничениях, «недостатках» индекса Хирша:

- h -индекс не предполагает поправки на дисциплинарную область: сравнение абсолютных значений индекса у ученых, работающих в разных областях науки, невозможно; это не нормализованный по областям науки показатель;

- h -индекс является целым числом, следовательно, не обладает идеальной «разрешающей способностью» (были предложены способы обобщить h -индекс на дробные значения, здесь мы не будем на них останавливаться);

- в индексе не учитывается реальное число ссылок, полученное статьями выше и ниже точки h : важно только местонахождение этой точки. Так, например, если бы каждая из h первых статей на рис. 11 получила в 10 раз больше ссылок (а цитируемость остальных публикаций осталась прежней), на значении h -индекса это бы ни-

¹⁹Строго говоря, все-таки существует нечастая ситуация, при которой h -индекс ученого способен уменьшиться. Это может произойти в случае, если ученый отозвал свою уже опубликованную в журнале статью (retraction) — из-за найденных серьезных ошибок или обнаружения плагиата и т. д. На это обстоятельство обращал внимание В. Глэнцел (в частной беседе).

как не отразилось. $(h + 1)$ -я статья все равно осталась бы лежать ниже линии $y = x$ и индекс Хирша сохранил свое значение. И наоборот: допустим, что два гипотетических автора опубликовали по 50 статей каждый. Каждая статья первого из них получила по 10 ссылок, а у второго 10 статей получили по 10 ссылок, а остальные 40 статей вообще не цитировались. В этом случае, при очевидном неравенстве «вклада» авторов, их индекс Хирша будет иметь одинаковое значение — 10.

Успех индекса Хирша был во многом связан с тем, что ряд ученых, а главное, администраторов науки, посчитали его простым и универсальным средством для оценки ученых/организаций и т. д. То, что эффективность научной деятельности можно охарактеризовать одним числом, оказалось слишком сильным соблазном. В действительности, как мы видим, h -индекс — это не панацея, а лишь один из библиометрических индикаторов, который должен использоваться обязательно в комплексе с другими метриками для разносторонней оценки авторов и научных коллективов.

2.6.2. h_α -индекс

Вообще говоря, условие равенства числа статей и минимального числа полученных каждой из этих статей ссылок является в определении индекса Хирша произвольным. Почему это не требование, чтобы h статей получили каждая не менее $2h$ ссылок? Или, наоборот, чтобы каждая из h статей получила не менее $h/2$ ссылок? Имея это в виду, Н. Ван Эйк и Л. Валтман [Van Eck, Waltman, 2008] «обобщили» определение индекса Хирша, назвав его h_α -индекс.

Определение: обобщенный h_α -индекс массива публикаций равен h_α , если каждая из h_α статей этого массива получила не ме-

нее ah цитирований, а каждая из остальных — не более ah цитирований.

Графически смысл h_α -индекса отражен на рис. 12. По графику h_α -индекс можно определить таким же образом, как и обычный индекс Хирша, но следует искать пересечение не с прямой $y = x$, а с прямой $y = \alpha x$. На рис. 12 показаны примеры для $\alpha = 2$ и $\alpha = 1/2$.

Основной смысл нового индекса в следующем. Как мы говорили, индекс Хирша одновременно оценивает число публикаций и их цитируемость, делая это в каком-то смысле «равномерно» (в смысле условия равенства числа статей и минимального числа полученных каждой из этих статей ссылок). При помощи h_α -индекса мы можем управлять этой «равномерностью». При увеличении α на первое место выходит цитируемость, наличие высокоцитируемых работ. При уменьшении — все важнее становится продуктивность, большое число публикаций. Это понятно из графического представления h_α -индекса, а мы продемонстрируем на упомянутом ранее примере: пусть один автор написал 5 статей, каждая из которых получила по 100 ссылок, а вто-

рой написал 100 статей, каждая из которых получила 5 ссылок. При $\alpha = 5$ у первого автора h_α -индекс равен 5 (есть 5 статей, каждая из которых процитирована не менее $\alpha \times 5 = 5 \times 5 = 25$ раз), а у второго — 1 (есть лишь 1 статья, которая процитирована не менее $\alpha \times 5 = 1 \times 5 = 5$ раз, а двух статей, получивших по 10 ссылок, нет). При $\alpha = 1/5$ у первого автора h_α -индекс остается равным 5, а у второго возрастает до 25 (есть 25 статей, каждая из которых процитирована не менее $\alpha \times 25 = (1/5) \times 25 = 5$ раз)¹⁶²⁰.

Таким образом, считая h_α -индекс при разных α , можно исследовать разные аспекты исследуемого множества статей, в зависимости от поставленных задач больше фокусируясь на публикационной активности, на числе публикаций, или делая основной акцент на качественном аспекте, на цитируемости изучаемых работ.

2.6.3. g -индекс

Еще одна попытка «улучшить» индекс Хирша — это введенный L. Egghe [Egghe, 2006] g -индекс. Он часто встречается в литературе (хотя и реже, чем h -индекс). По определению, g -индекс ученого равен такому максимальному g , что g наиболее цитируемых статей ученого получили вместе не менее g^2 ссылок. (Соответственно, любые $g + 1$ его статей совместно получили меньше $(g + 1)^2$ ссылок.) Обратите внимание, что здесь

²⁰Если рассмотреть предельные случаи, то для любого автора максимальное значение α , при котором его h_α -индекс все еще больше нуля (т. е. равен 1), совпадает с числом цитирований, которое получила его самая высокоцитируемая статья. Число статей при этом не играет никакой роли. Наоборот, при уменьшении α h_α -индекс возрастает и достигает своего максимального значения, которое равно общему числу статей данного автора, получивших хотя бы одну ссылку. Здесь совершенно не важна цитируемость статей (за исключением триггера «была ли статьей получена хотя бы одна ссылка»).

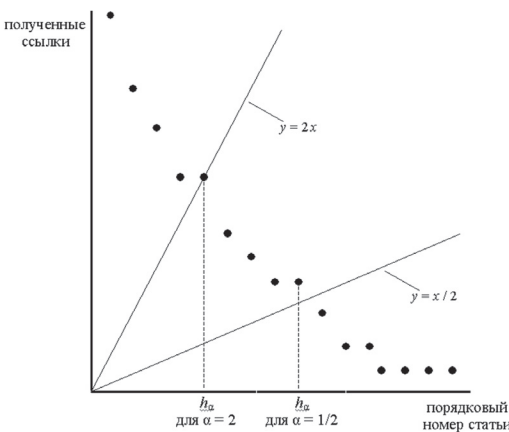


Рис. 12. h_α -индекс. Пример для $\alpha = 2$ и $\alpha = 1/2$

рассматривается совокупная цитируемость первых g статей, упорядоченных в порядке убывания цитируемости. В случае h -индекса требования по цитируемости предъявлялись к каждой отдельной статье, в этом ключевое отличие индексов друг от друга.

Вернемся к примеру из п. 2.6.1, где рассматривались пять работ ученого, которые получили 10, 8, 2, 2 и 0 ссылок соответственно. g -индекс этого автора равен 4 — поскольку есть 4 статьи, которые цитировались вместе не менее $4^2 = 16$ раз (на самом деле они цитировались $10 + 8 + 2 + 2 = 22$ раза), но нет 5 статей, которые бы в сумме цитировались $5^2 = 25$ раз. Напомним, h -индекс данного ученого оказался равным 2. В действительности g -индекс всегда больше или равен h -индексу. Это следует из того, что каждая из h статей получила, по определению индекса Хирша, не менее h ссылок, а значит, вместе они получили как минимум h^2 цитирований. Поэтому h статей заведомо удовлетворяют условию, которое предусмотрено g -индексом. Следовательно, $g \geq h$.

Можно переформулировать определение g -индекса так: g -индекс ученого равен такому максимальному g , что у ученого есть g статей со *средней цитируемостью* всей их совокупности (в расчете на одну статью) не менее g . Определения эквивалентны, поскольку наличие g статей со средней цитируемостью их совокупности не менее g означает, что в сумме эти статьи получили не менее g^2 ссылок.

В чем преимущество g -индекса? Он лишен одного из недостатков, отмеченных для индекса Хирша. Цитируемость наиболее цитируемых статей (получивших больше ссылок, чем статья с номером g) может влиять на значение g -индекса. Так, если в приведенном выше примере самая цитируемая статья со временем получит еще три ссыл-

ки (и их станет 13, цитируемость остальных статей пусть не меняется), g -индекс ученого достигнет пяти. Напомним, что в h -индексе не учитывается реальное число ссылок, полученное статьями выше и ниже точки h : важно только местонахождение этой точки. Любое увеличение цитируемости самой цитируемой статьи не повлияет на h -индекс.

2.7. Показатели хронологического распределения ссылок

В тексте данной главы неоднократно упоминалось о «старении» научного знания. Как быстро новое знание воспринимается, вводится в научный оборот и, напротив, как долго не теряет своей актуальности и остается востребованным. Эти характеристики можно оценить по хронологической структуре библиографических ссылок, которые содержатся в том или ином журнале, работах ученого, статьях института или, наоборот, которые цитируют данный журнал, работы данного ученого, статьи института. В данном случае речь уже не идет о попытке оценить качество, эффективность публикаций по числу полученных ими ссылок. Это другое исследование. И тем не менее журнал, цитирующий в основном классические источники многолетней давности, отличается от издания, стремящегося знакомить своих читателей с самой современной литературой по рассматриваемому вопросу. А если ученый опубликовал работу, в дальнейшем регулярно цитируемую из года в год, его вклад, скорее всего, будет иного характера, чем у исследователя, получившего много ссылок в первые 2–3 года после публикации статьи и впоследствии забытого.

Для хронологической характеристики распределения сделанных журналом/автором/

организацией цитирований чаще всего используют значение медианы этого распределения. Половина сделанных ссылок ведет на источники, опубликованные позже значения медианы, половина — на более свежую литературу. На рис. 13 показано хронологическое распределение ссылок, сделанных в статьях журнала *Annual Review of Biochemistry*, вышедших в 2013 г. В этих статьях была 21 ссылка на публикации 2013 г., 424 ссылки на источники, вышедшие в 2012 г., и т. д. В общей сложности статьи 2013 г. содержали 4883 ссылки, из них половина пришлась на документы, вышедшие в 2007–2013 гг., половина — на более поздние источники. Поэтому медиана соответствует документам семилетней давности. В JCR для этого показателя используют термин «время полужизни сделанных ссылок» (citing half-life). Таким образом, для данного журнала в 2013 г. время полужизни сделанных ссылок равно 7 (затенен-

ной областью на рисунке отмечены ссылки на более поздние источники)¹⁷²¹.

Время полужизни ссылок зависит от научной области, к которой относится журнал. Собственно, оно и характеризует эту область с точки зрения темпов устаревания производимого в ней научного знания. В JCR-2013 минимальное время полужизни сделанных ссылок у журналов дисциплины «нанонауки и нанотехнологии» (5,9 лет). Максимальный уровень этого показателя оценить сложно, поскольку JCR не указывает его точное значение для тех областей науки, где оно больше 10 лет. Таких дисциплин в JCR-2013 в общей сложности 48, среди них, в частности, история, социология, зоология, геология, лингвистика и др. В этих областях научное знание имеет максимально продолжительный цикл жизни, и больше половины всех публикуемых ссылок ведет на источники более чем десятилетней давности.

Аналогичный показатель также вводится не для сделанных ссылок, а, наоборот, для полученных журналом цитирований. Он называется *cited half-life*, время полужизни полученных ссылок, и характеризует то, насколько свежие статьи журнала цитируются в мировой науке. Если его величина мала, значит, в науке используются прежде всего свежие, недавние публикации издания. Если велика — полезными оказываются и достаточно старые материалы, публиковавшиеся в журнале. Время полужизни полученных ссылок для целых научных дисциплин варьируется в JCR-2013 от 4,1 года (снова нанонауки) до «более 10 лет» (31 дисциплина). Безусловно, *cited half-life* связан с *citing*

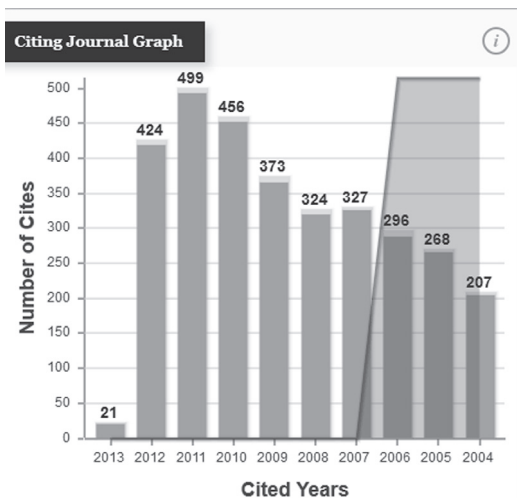


Рис. 13. Хронологическое распределение ссылок в статьях журнала *Annual Review of Biochemistry*, вышедших в 2013 г. Скриншот интерфейса базы данных JCR (Thomson Reuters)

²¹Если быть точным, в приведенном примере на публикации 2007–2013 гг. приходится 49,6 % всех ссылок (2424 из 4883). Если добавить еще один год (2006), то будет охвачено уже 55,7 % всех ссылок. Поэтому делается пропорциональная поправка и время полужизни для этого издания указывается в JCR равным 7,1 года.

half-life: напрямую они не зависят друг от друга, однако оба показателя отражают скорость эволюции знания в соответствующей научной области. Поэтому, например, среди 31 дисциплины со временем полужизни полученных ссылок более 10 лет в 28 дисциплинах время полужизни сделанных ссылок также превышает 10 лет.

Оба изученных индикатора хронологического распределения ссылок в журналах с легкостью могут быть распространены на статьи, написанные отдельным ученым или каким-либо научным коллективом.

ЗАКЛЮЧЕНИЕ

Мы рассмотрели основные библиометрические показатели, характеризующие цитируемость разных «научных единиц» — ученых, научных коллективов, организаций и целых стран. Были также разобраны разнообразны журнальные показатели. Заметим, что это далеко не исчерпывающий список индикаторов, существующих в библиометрии, и даже не исчерпывающий перечень индикаторов, доступных в инструментах Thomson Reuters. Не имеющие прямого отношения к цитируемости и оценке эффективности научной деятельности, нами были оставлены в стороне такие вопросы, как измерение международного сотрудничества в публикациях, характеристики сетей соавторства, индексы научной специализации, коэффициенты ко-цитирования и др.

Иные показатели не рассматривались ввиду их тривиальности — например, интегральная публикационная активность или суммарная цитируемость того или иного объекта исследования.

Неоднократно в течение настоящей главы встречалось напоминание о том, что идеальных индикаторов не существует, что они должны использоваться не по отдельности, а в комплексе. Цель, поставленная здесь нами, — научить грамотно и многосторонне изучать эффективность науки с помощью библиометрии. Но результаты даже самого подробного библиометрического исследования не отменяют необходимости участия экспертов, профессионалов-предметников при принятии решений из области science policy: при распределении грантов и финансировании, осуществлении кадровой политики, открытии/закрытии научных направлений, программ и лабораторий и т. д. На основании чистой библиометрии не могут выноситься решения на поле управления наукой — ее роль заключается в том, чтобы помогать в принятии таких решений, сообщать дополнительные сведения об изучаемых «научных единицах»: ученых, лабораториях, факультетах, организациях, регионах, странах, журналах. В западном наукометрическом сообществе в таких случаях говорят о концепции «informed decision», «решения при полной информации», в реализации которой и должен играть свою роль библиометрический анализ.

Писляков В. В. Библиометрические индикаторы: практикум. М.: НФПК; Инфра-М, 2014. 60 с.

Писляков В. В. Соавторство российских ученых с зарубежными коллегами: публикации и их цитируемость. Препринт WP6/2010/01. М.: ГУ-ВШЭ, 2010. 40 с.

Писляков В. В. Шедевры научного творчества: анализ высокоцитируемых статей российских ученых // На-

учно-техническая информация. Сер. 2: Информационные процессы и системы. 2011. № 12. С. 1–8.

Писляков В. В., Дьяченко Е. Л. Эффект Матфея в цитировании статей российских ученых, опубликованных за рубежом // Научно-техническая информация.

Сер. 2: Информационные процессы и системы. 2009. № 3. С. 19–24.

Aksnes D. W., Sivertsen G. The effect of highly cited papers on national citation indicators. *Scientometrics*. 2004. Vol. 59. Iss. 2. P. 213–224.

Eck N. J. van, Waltman L. Generalizing the h- and g-indices // *Journal of Informetrics*. 2008. Vol. 2. Iss. 4. P. 263–271.

Egghe L. Theory and practise of the g-index // *Scientometrics*. 2006. Vol. 69. Iss. 1. P. 131–152.

Hirsch J. E. An index to quantify an individual's scientific research output // *Proceedings of the National Academy of Sciences*. 2005. Vol. 102. № 46. P. 16569–16572.

Larivière V., Archambault É., Gingras Y., Vignola-Gagné É. The place of serials in referencing practices: Comparing natural sciences and engineering with social sciences and humanities // *Journal of the American Society for Information Science and Technology*. 2006. Vol. 57. Iss. 8. P. 997–1004.

McVeigh M. E. Journal self-citation in the Journal Citation Reports – Science Edition (2002). 2004. URL: <http://wokinfo.com/essays/journal-self-citation-jcr>.

Pislyakov V., Dyachenko E. Citation expectations: are they realized? Study of the Matthew index for Russian papers published abroad // *Scientometrics*. 2010. Vol. 83. № 3. P. 739–749.

Rousseau R. Journal evaluation: Technical and practical issues // *Library Trends*. 2002. Vol. 50. Iss. 3. P. 418–439.

Помимо справочного аппарата данной главы, для более глубокого знакомства с библиометрическими индикаторами читателю рекомендуются следующие отечественные и зарубежные источники:

Бредихин С. В., Кузнецов А. Ю., Щербакова Н. Г. Анализ цитирования в библиометрии. Новосибирск; М.: ИВМиМГ СО РАН, НЭИКОН, 2013.

Кузнецов А. Ю., Бредихин С. В. Методы библиометрии и рынок электронной научной периодики. Новосибирск; М.: ИВМиМГ СО РАН, НЭИКОН, 2012.

Писляков В. В. Методы оценки научного знания по показателям цитирования // *Социологический журнал*. 2007. № 1. С. 128–140.

Egghe L., Rousseau R. *Introduction to Informetrics: Quantitative Methods in Library, Documentation and Information Science*. Amsterdam e. a.: Elsevier Science Publishers, 1990.

Garfield E. Citation indexes for science. A new dimension in documentation through association of ideas // *Science*. 1955. Vol. 122. № 3159. P. 108–111.

Garfield E., Sher I. H. New factors in the evaluation of scientific literature through citation indexing // *American Documentation*. 1963. Vol. 14. № 3. P. 195–201.

Glänzel W. *Bibliometrics as a Research Field: A course on theory and application of bibliometric indicators*. Course Handouts. Leuven, 2003.

Glänzel W., Moed H. F. Journal impact measures in bibliometric research // *Scientometrics*. 2002. Vol. 53. Iss. 2. P. 171–193.

Gonzalez-Pereira B., Guerrero-Bote V., Moya-Anegón F. A new approach to the metric of journals scientific prestige: The SJR indicator // *Journal of Informetrics*. 2010. Vol. 4. Iss. 3. P. 379–391.

Handbook of Quantitative Science and Technology Research / ed. by H. F. Moed, W. Glänzel, U. Schmoch. Dordrecht: Kluwer Academic Publishers, 2004.

Moed H. F. *Citation Analysis in Research Evaluation*. Dordrecht: Springer, 2005.

Moed H. F. Measuring contextual citation impact of scientific journals // *Journal of Informetrics*. 2010. Vol. 4. Iss. 3. P. 265–277.

Pislyakov V. Comparing two «thermometers»: Impact factors of 20 leading economic journals according to Journal Citation Reports and Scopus // *Scientometrics*. 2009. Vol. 79. № 3. P. 541–550.

Pislyakov V., Shukshina E. Measuring excellence in Russia: Highly cited papers, leading institutions, patterns of national and international collaboration // *Journal of the Association for Information Science and Technology*. 2014. Vol. 65. Iss. 11. P. 2321–2330.

Rehn C., Gornitzki C., Larsson A., Wadskog D. *Bibliometric Handbook for Karolinska Institutet*. 2014. URL: https://kib.ki.se/sites/kib.ki.se/files/Bibliometric_Handbook_2014.pdf.

Thomson Reuters. *InCites Indicators Handbook*. 2014. URL: <http://researchanalytics.thomsonreuters.com/m/pdfs/indicators-handbook.pdf>.