

БИБЛИОМЕТРИЧЕСКИЕ ИНДИКАТОРЫ В РЕСУРСАХ КОМПАНИИ CLARIVATE*



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

BIBLIOMETRIC INDICATORS IN THE CLARIVATE DATABASES

DOI 10.15826/B978-5-7996-3154-3.008

This chapter examines bibliometric indicators related to citedness of journals, authors, research groups, organizations and whole countries. The author discusses usage of various bibliometric indicators: the impact factor, average citedness, share of uncited papers, Eigenfactor and Article Influence Score, 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 and by publication sources, as well as to creation of normalized citation profiles. The introductory section deals with the basics of bibliometric analysis and features of citation databases. The final section emphasizes importance of informed and responsible use of bibliometric indicators in research policy-making, funding allocation, and faculty and research personnel recruitment.

Keywords: *bibliometric indicators, Web of Science, normalized citedness, Hirsch index, highly cited papers.*

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. Особное внимание уделено нормализованным индикаторам, с нормализацией по областям науки и по журналам, а также построению профилей нормализованной цитируемости. Глава содержит вводную часть, в которой излагаются основы библиометрического анализа и характеристики баз данных научного цитирования. Финальный раздел, заключение, подчеркивает необходимость грамотной и аккуратной трактовки библиометрических индикаторов при принятии административных решений, распределении грантов, осуществлении кадровой политики и т. д.

Ключевые слова: *библиометрические индикаторы, Web of Science, нормализованная цитируемость, индекс Хирша, высокоцитируемые статьи.*

*Глава написана при поддержке Российского научного фонда, проект № 20-18-00140 «Структура и траектории науки на постсоветском пространстве: институциональные практики и библиометрический анализ».

Введение

В данной главе мы рассмотрим ряд библиометрических индикаторов, которые либо содержатся в готовом виде, либо могут быть рассчитаны при помощи баз данных научного цитирования компании Clarivate. Все эти индикаторы связаны с *цитируемостью* научных публикаций¹⁹.

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 вступном разделе 3.1 рассказывается, что содержится в базах данных научного цитирования, каково наполнение и функции инструментов, предоставляемых компанией Clarivate, какие существуют обозначения, терминологические и методологические договоренности в литературе по библиометрии. Раздел 3.2 посвящен импакт-индикаторам, измеряющим среднюю цитируемость в расчете на одну статью; здесь же затрагивается вопрос о самоцитировании в научной литературе. В разделе 3.3 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 3.4 вводится понятие

о показателях «экстремальной» (предельной) цитируемости, к которым отнесены доля высокоцитируемых статей и коэффициент нецитируемости. Раздел 3.5 — совершенно новый, написанный специально для второго издания «Руководства...». В нем объясняется одна из наиболее перспективных методик на данный момент — построение «профилей цитируемости» авторов и организаций. «Взвешенные» индикаторы, учитывающие при подсчете цитируемости научный уровень цитирующих журналов, освещаются в разделе 3.6. Индексу Хирша и некоторым его модификациям посвящен раздел 3.7. В разделе 3.8 исследуются показатели, характеризующие хронологическую структуру распределения ссылок, и, наконец, Заключение завершает данную главу.

3.1. Библиометрические инструменты.

База данных Web of Science Core Collection и аналитические надстройки

Начнем с описания инструмента Web of Science Core Collection (Web of Science CC), продукта компании Clarivate. 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, по массивам исходных («сырых») научных данных, ряд те-

¹⁹ Необходимо заметить, что английское «citation», как справедливо указывалось, например, А. В. Полетаевым или Р. С. Гиляревским, означает не «цитирование» или тем более «цитату», а 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.

матических реферативных баз, национальные (страновые) базы научного цитирования, в том числе Russian Science Citation Index (RSCI)²⁰, и др.

3.1.1. 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:

- 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 имела в исходном тексте публикации; базы цитирования не составляют аннотации для тех произведений, в которых они изначально отсутствуют;
- ключевые слова; иногда это два набора ключевых слов — слова, приписанные статье в оригинальном тексте публикации («авторские ключевые слова»); и ключевые слова, «назначенные» базой данных цитирования исходя из ее внутреннего тезауруса ключевых слов и автоматических алгоритмов (они анализируют заголовки цитируемых в статье других источников);
- тематика (рубрика), приписанная статье, и тип публикации (см. далее);
- организации, в которых работают авторы, так называемые «аффилиации», с почтовыми 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, информация о грантовой поддержке, название и адрес издательства и др.

Таким образом, если журнал «расписывается» (индексируется) библиометрической базой данных, это означает, что о каждой публикации данного журнала мы сможем узнать из базы данных сведения, перечисленные выше. Полного текста статей при этом в базе не будет. В качестве иллюстрации на рис. 33 приведен пример библиографической записи Web of Science CC на некоторую журнальную статью, а на рис. 34 — список литературы этой статьи (он открывается при нажатии соответствующей ссылки).

Помимо журнального контента, библиометрические базы данных могут включать в себя также труды конференций и книги. При этом содержание полей, попадающих 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 име-

²⁰ Подробнее об RSCI см. 2.2 с. 151.

Strain-hardening and suppression of shear-banding in rejuvenated bulk metallic glass

Автор: Pan, J (Pan, J.)^[1]; Ivanov, YP (Ivanov, Yu. P.)^[2,3]; Zhou, WH (Zhou, W. H.)^[1,4]; Li, Y (Li, Y.)^[1]; Greer, AL (Greer, A. L.)^[2]

Показать номер Web of Science ResearcherID и ORCID

NATURE

Том: 578 Выпуск: 7796 Стр.: 559+

DOI: 10.1038/s41586-020-2016-3

Опубликовано: FEB 2020

Тип документа: Article

Просмотреть Impact Factor журнала

Аннотация

Strain-hardening (the increase of flow stress with plastic strain) is the most important phenomenon in the mechanical behaviour of engineering alloys because it ensures that flow is delocalized, enhances tensile ductility and inhibits catastrophic mechanical failure^(1,2). Metallic glasses (MGs) lack the crystallinity of conventional engineering alloys, and some of their properties—such as higher yield stress and elastic strain limit⁽³⁾—are greatly improved relative to their crystalline counterparts. MGs can have high fracture toughness and have the highest known 'damage tolerance' (defined as the product of yield stress and fracture toughness)⁽⁴⁾ among all structural materials. However, the use of MGs in structural applications is largely limited by the fact that they show strain-softening instead of strain-hardening; this leads to extreme localization of plastic flow in shear bands, and is associated with early catastrophic failure in tension. Although rejuvenation of an MG (raising its energy to values that are typical of glass formation at a higher cooling rate) lowers its yield stress, which might enable strain-hardening⁽⁵⁾, it is unclear whether sufficient rejuvenation can be achieved in bulk samples while retaining their glassy structure. Here we show that plastic deformation under triaxial compression at room temperature can rejuvenate bulk MG samples sufficiently to enable strain-hardening through a mechanism that has not been previously observed in the metallic state. This transformed behaviour suppresses shear-banding in bulk samples in normal uniaxial (tensile or compressive) tests, prevents catastrophic failure and leads to higher ultimate flow stress. The rejuvenated MGs are stable at room temperature and show exceptionally efficient strain-hardening, greatly increasing their potential use in structural applications.

Bulk metallic glasses can acquire the ability to strain-harden through a mechanical rejuvenation treatment at room temperature that retains their non-crystalline structure.

Ключевые слова

KeyWords Plus: TENSILE DUCTILITY; FRACTURE; MECHANISMS; TRANSITION; NECKING; STATE

Информация об авторе

Адрес для корреспонденции: Li, Y (автор для корреспонденции)

+ Chinese Acad Sci, Shenyang Natl Lab Mat Sci, Inst Met Res, Shenyang, Peoples R China.

Адрес для корреспонденции: Greer, AL (автор для корреспонденции)

+ Univ Cambridge, Dept Mat Sci & Met, Cambridge, England.

Адреса:

+ [1] Chinese Acad Sci, Shenyang Natl Lab Mat Sci, Inst Met Res, Shenyang, Peoples R China

+ [2] Univ Cambridge, Dept Mat Sci & Met, Cambridge, England

+ [3] Far Eastern Fed Univ, Sch Nat Sci, Vladivostok, Russia

+ [4] Univ Sci & Technol China, Sch Mat Sci & Engrg, Hefei, Peoples R China

Адреса эл. почты: lly1@imr.ac.cn; alg13@cam.ac.uk

Рис. 33. Запись в библиометрической базе данных на статью из журнала Nature. Скриншот интерфейса базы данных Web of Science CC

ется 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).

Пристатейных ссылок: 43

(из Web of Science Core Collection)

ИЗ: Strain-hardening and suppression of shear-banding in rejuvenated bulk metallic glass ...**Больше**

Добавьте в список отмеченных публикаций

1. **Metallic glasses as structural materials**
Автор: Ashby, MF; Greer, AL
SCRIPTA MATERIALIA Том: 54 Выпуск: 3 Стр.: 321-326 Опубликовано: FEB 2006
 [Просмотреть аннотацию ▼](#)
2. **SERIALLY DEPOSITED AMORPHOUS AGGREGATES OF HARD SPHERES**
Автор: BENNETT, CH
JOURNAL OF APPLIED PHYSICS Том: 43 Выпуск: 6 Стр.: 2727-& Опубликовано: 1972

3. Название: [нет доступа]
Автор: Boyer, H. E.
Atlas of Stress-Strain Curves Опубликовано: 2002
Издатель: ASM International

4. **Ductility and work hardening in nano-sized metallic glasses**
Автор: Chen, D. Z.; Gu, X. W.; An, Q.; с соавторами.
APPLIED PHYSICS LETTERS Том: 106 Выпуск: 6 Номер статьи: 061903 Опубликовано: FEB 9 2015
 [Просмотреть аннотацию ▼](#)

Рис. 34. Список цитируемой в статье литературы.
Скриншот интерфейса базы данных Web of Science CC

В базах научного цитирования каждый журнал имеет «тематическую привязку», он отнесен к той или иной научной дисциплине (может быть отнесен более чем к одной). Что важно: 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*, где она опубликована. Все статьи одного журнала имеют одну и ту же рубрику (рубрики). Мы будем говорить, что

статья/журнал относятся к той или иной тематической рубрике, или тематической категории, или дисциплине, или научной области. Все эти понятия используются нами как синонимы.

Помимо дисциплины, каждой публикации в библиометрических базах данных присваивается *тип документа*. Это может быть научная исследовательская 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». Оговоримся, что словом «статья» мы часто будем обозначать, как это делается в обыденной практике, любую публикацию в журнале. В тех разделах, где будет важно отличать произвольную «публикацию» от «статьи» (то есть 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 ученый 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, ко-

нечно, 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 подсчета 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. Проводить дробные расчеты на практике — очень трудоемкое занятие, однако в последнее время все чаще считается, что именно такой учет дает более взвешенное и справедливое представление о вкладе ученого или организации в публикации. Так, предложенный в начале 2020 г. Министерством науки и высшего образования РФ индикатор КБПР (комплекс-

ный балл публикационной результативности) использует именно fractional counting, дробный счет.

Тем не менее во всех индикаторах, которые рассматриваются 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.

3.1.2. *Web of Science Core Collection*

Итак, Web of Science CC является базой данных цитирования, «библиометрической» базой данных. Она содержит восемь блоков (или «изданий», editions) и в сумме охватывала на начало 2020 г. более 21000 научных журналов, более 210000 сборников трудов конференций и более 110000 книг. Web of Science CC является ядром библиометрических продуктов Clarivate, 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 1900 г.;

- Social Sciences Citation Index (SSCI) — журналы по общественным наукам (экономика, менеджмент, социология, право, политология, история и т. д.); архив с 1900 г.;

- Arts & Humanities Citation Index (A&HCI) — журналы по гуманитарным наукам (история, литературоведение, искусствоведение, религиоведение и т. д.); архив с 1975 г.;

- Emerging Sources Citation Index — журналы всего спектра научных дисциплин, не вошедшие в указанные выше индексы (часто называемые «старшими индексами»), но оцениваемые Clarivate на предмет включения в них; своего рода «приемная» WoS CC, содержащая много более локальных и менее известных изданий; архив с 2005 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 по естественным, техническим, медицинским наукам; архив с 2005 г.;

- Book Citation Index — Social Sciences & Humanities — книги по общественным и гуманитарным наукам; архив с 2005 г.

Кроме этого, в WoS CC входят две базы данных по химическим соединениям и реакциям (Index Chemicus и Current Chemical

Reactions), однако мы никак не будем их касаться.

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

В настоящее время Web of Science CC индексирует 148 российских изданий в SCIE, 3 журнала в SSCI и 10 — в A&HCI.

3.1.3. 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 готовые 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. 3.2.1) 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. 3.4.1), среднемировым библиометрическим показателям («*baselines*») и перспективным научным «фронтам», которые определяются специальной библиометрической процедурой.

InCites (*InCites Benchmarking & Analytics*) — 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 (см. раздел 3.3).

В JCR и ESI показатели рассчитываются только для естественно-научных и общественно-научных журналов, то есть для баз SCIE и SSCI. В InCites учитываются все журнальные, книжные и конференционные ресурсы WoS CC.

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 использо-

вания в специальных аналитических «надстройках» JCR, ESI и InCites.

3.2. 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 2014–2018 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 2018 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 примере мы бы считали среднее число ссылок, присутствующих в статьях 2018 г. выхода, на работы, опубликованные в 2014–2018 гг. учеными,

работающими в вузе U . «Среднее» обозначает здесь усреднение по числу оцениваемых статей: найденное суммарное число ссылок должно быть поделено на число статей в публикационном окне (то есть на число статей, опубликованных учеными вуза U в 2014–2018 гг.).

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

3.2.1. 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 Y статьями данного журнала, вышедшими в годах $Y-1$ и $Y-2$, к числу этих статей (то есть 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).

Чтобы пояснить еще более наглядно, введем обозначения. Переменная P_x будет обозначать число статей, вышедших в оцениваемом журнале в год x , а $C_{y \rightarrow x}$ число ссылок, полученных в году y статьями данного жур-

Пример

В 2018 г. во всех источниках, охваченных Web of Science CC, было сделано 189 ссылок на статьи журнала «Успехи физических наук», опубликованные в 2017 г., и 259 ссылок на его статьи, опубликованные в 2016 г. Итого науки «Успехи...», вышедшие в 2016–2017 гг., получили в 2018 г. $189 + 259 = 448$ цитирований. При этом в 2017 г. журнал опубликовал 72 статьи, в 2016 г. — 73, итого за промежутки 2016–2017 гг. («публикационное окно») вышло 145 статей. Таким образом, импакт-фактор журнала «УФН» в 2018 г. равен отношению $448 : 145 = 3,09$.

нала, вышедшими в год x . Тогда формула для импакт-фактора IF журнала в 2018 г. будет иметь вид:

$$IF = \frac{C_{2018 \rightarrow 2017} + C_{2018 \rightarrow 2016}}{P_{2017} + P_{2016}}.$$

Резюмируя, импакт-фактор характеризует среднее число ссылок, полученных в отчетном году статьями журнала, опубликованными в течение двух предыдущих лет²¹. См. иллюстрацию этого конкретным примером.

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 (то есть импакт-факторы 2019 г. можно узнать летом 2020 г. и т. д.). На рис. 35 приведен скриншот базы данных JCR с журналами по акустике, которые упорядочены по убыванию импакт-фактора. Кроме того, многие издательства размещают текущие значения импакт-факторов своих журналов на их веб-страницах, в открытом доступе. Нередко они сопровож-

²¹ Нюансы, которые необязательно запоминать при первом знакомстве с импакт-фактором. В числителе сейчас учитываются ссылки не только из всех журналов WoS CC (включая ESCI), но и из книжных и конференционных индексов. При этом считаются цитирования любой публикации журнала, вне зависимости от ее типа. В то же время в знаменателе подсчитываются только, как их называют, «документы, которые могут быть процитированы» (citable items). Это упомянутые ранее исследовательские статьи (Article) и научные обзоры (Review). Такое различие в подходе к типам документа числителя и знаменателя иногда считается слабым местом методики расчета импакт-фактора.

	Full Journal Title	Total Cites	Journal Impact Factor ▼
1	ULTRASONICS SONOCHEMISTRY	17,314	7.279
2	ULTRASOUND IN OBSTETRICS & GYNECOLOGY	12,336	5.595
3	ULTRASCHALL IN DER MEDIZIN	2,238	4.613
4	IEEE-ACM Transactions on Audio Speech and Language Processing	3,110	3.531
5	JOURNAL OF SOUND AND VIBRATION	36,167	3.123
6	IEEE TRANSACTIONS ON ULTRASONICS FERROELECTRICS AND FREQUENCY CONTROL	11,266	2.989
7	JOURNAL OF VIBRATION AND CONTROL	5,075	2.865
8	ULTRASONICS	7,026	2.598

Рис. 35. Журналы по акустике, в порядке убывания импакт-фактора.

Скриншот интерфейса базы данных JCR (Clarivate)

ждаются ранговыми показателями (см. далее, п. 3.3.4).

На данный момент (первая половина 2020 г., последнее вышедшее издание — JCR-2018) максимальный импакт-фактор имеет журнал «CA — A Cancer Journal for

Clinicians», IF = 223,7. Это совершенно невероятный показатель (журнал публикует малое число статей, но получает очень много ссылок), и ближайший «конкурент», журнал Nature Reviews Materials, отстает от него ровно в три раза, IF = 74,4. 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 журналов, входящих в SCIE (то есть журналов по естественным, техническим, медицинским наукам), равна 1,91. То есть среди всех 9153 журналов с импакт-фактором половина имеет импакт выше 1,91, а половина — ниже. Для общественно-научных журналов SSCI медиана еще ниже, 1,40.

Обратим внимание: JCR не публикует импакт-факторы гуманитарных журналов (то есть 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, Clarivate следует политике вообще не публиковать импакт-факторы для изданий по гуманитарным наукам.

Также не рассчитывается импакт-фактор журналов из ESCI ввиду особого «предва-

рительного» статуса этой базы данных. При этом, как сказано выше, ссылки из A&HCI и ESCI учитываются при подсчете числителя импакт-фактора журналов из SCIE и SSCI.

3.2.2. 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. 3.2.1. Он учитывает цитирования, которые статьи журнала получают в течение двух лет после своего выхода. Однако в ряде научных дисциплин, особенно в социальных 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 (то есть 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. В введенных нами обозначениях формула пятилетнего импакта за 2018 г. будет выглядеть так:

$$IF_5 = \frac{C_{2018 \rightarrow 2017} + C_{2018 \rightarrow 2016} + C_{2018 \rightarrow 2015} + C_{2018 \rightarrow 2014} + C_{2018 \rightarrow 2013}}{P_{2017} + P_{2016} + P_{2015} + P_{2014} + P_{2013}}$$

Если упорядочить журналы по убыванию пятилетнего импакт-фактора, результат будет отличаться от ранжирования по обычному импакту: журналы, чьи материалы устаревают медленнее, то есть 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 3.8).

Максимальный пятилетний импакт-фактор для журналов из SCIE равен 177,3, медиана — 1,99. Она несколько выше медианы двухлетнего импакт-фактора (1,91), это показывает, 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 производственного цикла научных издательств. Показатель, фиксирующий цитирования «того же года», также публикуется в базе данных JCR и называется immediacy index (II). Будем называть его «индексом оперативности» (дру-

гой вариант перевода, также предлагавшийся в литературе,— «индекс немедленного цитирования»).

Индекс оперативности предполагает однолетнее публикационное окно и однолетнее 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. В уже привычных обозначениях, для индекса 2018 года:

$$II = \frac{C_{2018 \rightarrow 2018}}{P_{2018}}.$$

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 II в JCR-2018 равен 52,6 (у того же журнала «CA—A Cancer Journal for Clinicians»). Медиана по журналам из SCIE предсказуемо мала — 0,49. Практически для каждого журнала индекс оперативности — самый низкий показатель из трех рассмотренных импакт-индикаторов (однако встречаются и исключения, например Physics of Life Reviews или American Journal of Bioethics, где индекс оперативности, наоборот, выше двух- и пятилетнего импактов).

В заключение рассказа о базовых журнальных показателях замечание о динамике изученных индикаторов. Если вы посмотрите на средние и максимальные показатели, приведенные в первом издании данной книги (2014), то обнаружите, что за прошедшее время они существенно подросли. Например, медиана двухлетнего импакт-фактора по SCIE выросла с 1,41 до 1,91 — больше, чем на треть. Журнал с макси-

мальным импактом поднялся от огромного значения $IF = 162,5$ до совсем невероятного $223,7$. Даже медиана самого маленького показателя, индекса оперативности, выросла с $0,27$ до $0,49$, плюс 80% . Вероятно, это вызвано добавлением коллекции ESCI в Web of Science, но и до ее включения прослеживалась общемировая тенденция к росту импакт-индикаторов. С чем это связано — вопрос, требующий специального исследования (например, скорее всего постепенно увеличивается средний размер списка литературы в научных статьях), однако эмпирический факт несомненен. Отсюда следует важный практический вывод. Если у вашего (или какого-то другого) журнала в новом году вырос импакт-фактор, это совсем не обязательно означает, что журнал стал лучше, что это некий *прогресс*. Тем более неправы те издательства, которые часто с выходом нового года JCR публикуют позитивные отчеты со списками своих журналов, где вырос импакт. Это может быть следствием общей тенденции, а может быть даже отставанием от нее. Единственно корректный подход здесь — сравнение с изменениями у других аналогичных изданий, сопоставление себя с общим фоном журналов той же научной дисциплины. Например, возможно использование рангового метода, о котором мы будем говорить далее в разделах 3.3.4 и 3.5.

3.2.3. 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 характеризующее 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 (среднее в расчете на одну статью).

В ESI такая 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, телекоммуникации, физическая химия), 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 «Citation Impact» (так обозначена цитируемость в расчете на одну статью) на любом публикационном окне, но окно цитирования всегда будет «по настоящее время».

3.2.4. Совокупный и средневзвешенный импакт-фактор

Как следует из определения, импакт-фактор является не характеристикой журнала как целого, а показателем средней влия-

тельности, 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, характеризует ее подлинную влияние, а импакт-фактор издания, в котором она опубликована, делает это лишь косвенно и недостоверно. Поэтому в п. 3.2.3 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. 3.2.3 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 им-

пакт-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 определен выше, к общему числу опубликованных организацией/ученым 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 3.3. Если же ограничиться рамками одной дисциплины (например, исследовать деятельность отдельного ученого или узкотематической лаборатории), то полученное значение средневзвешенного импакта интересно сравнить с им-

Пример

Посчитаем совокупный и средневзвешенный импакт-фактор журналов, в которых публиковались сотрудники Института астрономии РАН в прошедшей части 2020 г. (данные на середину апреля). Всего к текущему моменту опубликовано 18 статей, приведем распределение их по журналам:

Monthly Notices of the Royal Astronomical Society (7 статей), IF = 5,23

Astronomy Reports (3), IF = 1,24

Nature Astronomy (2), IF = 10,50

Astronomy & Astrophysics (1), IF = 6,21

Astroparticle Physics (1), IF = 2,60

Astrophysical Journal (1), IF = 5,58

Astrophysics (1), IF = 0,64

Nuclear Instruments Methods in Physics Research Section A Accelerators Spectrometers Detectors and Associated Equipment (1)

Research in Astronomy and Astrophysics (1), IF = 1,25

Импакт-фактор указан за последний доступный год (JCR-2018). Один из журналов (Nuclear Instruments...) не относится к тематической категории WoS «Astronomy & Astrophysics», поэтому статью, опубликованную в нем, мы учитывать не будем: как мы помним, средние и совокупные импакт-факторы можно считать только в рамках одной дисциплинарной области. Для оставшихся 17 статей совокупный импакт-фактор равен:

$$5,23 \times 7 + 1,24 \times 3 + 10,5 \times 2 + 6,21 \times 1 + 2,6 \times 1 + 5,58 \times 1 + 0,64 \times 1 + 1,25 \times 1 = \mathbf{77,61}$$

Средневзвешенный импакт-фактор, соответственно, будет равен $77,61 : 17 = \mathbf{4,57}$. Много это или мало для астрономического института? Если бы существовал журнал по астрономии с точно таким импакт-фактором, в рейтинге категории «Astronomy & Astrophysics» он бы находился между 16-м и 17-м местом из 69 журналов, отнесенных в JCR-2018 к этой рубрике. Это граница «первого квартала», первой четверти списка рейтинга. Вкратце можно сказать, что публикационный поток Института астрономии на старте 2020 г. достаточно высокого качества, хотя и не все сотрудники публикуются в наиболее высокоимпактных, лидирующих журналах.

В качестве небольшого комментария (не имеющего уже отношения к совокупному и средневзвешенному импактам) добавим, что журнал с наибольшим импакт-фактором, где опубликованы статьи сотрудников института в 2020 г., Nature Astronomy, занимает четвертое место в рейтинге «Astronomy & Astrophysics». Журнал, где опубликовано максимальное число статей, Monthly Notices of the Royal Astronomical Society, расположен на 15-й строчке в рубрике.

пакт-фактором всех других журналов данной дисциплины. Так можно определить, в «сильных» или «слабых» журналах публикуется в среднем 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 дисциплины.

Заметим, что, как упоминалось в п. 3.2.1, распределение импакт-факторов часто 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. В таких слу-

чаях более корректным показателем может быть медианный импакт-фактор журналов, в которых публикуется ученый/лаборатория, вместо среднего (средневзвешенного) импакта. Соответственно, сравнивать его следует с медианным значением по всей дисциплине²².

Благодаря использованию импакт-факторов журналов мы смогли в приведенном в тексте примере оперативно, в реальном времени, оценить текущее качество публикационного потока Института астрономии РАН. В этом главное преимущество использования данного метода. Чтобы получить реальное число ссылок, пришлось бы ждать 2–5 лет, в зависимости от дисциплины. Дискуссия о валидности обращения к импактам как способу дать оценку качеству опубликованных статей не прекращается в наукометрическом и научном сообществе. В конечном итоге это философский вопрос о соотношении части и целого. Понятно, что высокий импакт-фактор не гарантирует большое число ссылок каждой статье журнала. Но и говорить о том, что эти характеристики никак не связаны, по меньшей мере, странно. Заинтересованный читатель может ознакомиться с работами на эту тему, например [Seglen, 1997; Callaway, 2016; Yuret, 2016; Grančay et al., 2017; Tregoning, 2018].

3.2.5. 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, 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

²² Следует отметить, что приведенный выше пример является редким исключением: медианный импакт-фактор публикаций Института астрономии РАН (5,23) выше среднего (4,57). Здесь отдельные статьи в слабых журналах, наоборот, занижают среднее значение по всему публикационному потоку.

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. Иногда здесь можно предположить наличие нездоровой 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, хорошо известных и цитируемых широким кругом источников. Например, у российских изданий «Светотехника» и «Новые горизонты», согласно JCR-2018, коэффициент самоцитируемости составляет 49 и 47% соответственно²³. Оба журнала находятся на втором снизу месте в импакт-рейтингах своих дисциплин. Если взять журналы-лидеры в соответствующих рубриках JCR-2018, то для них коэффициент самоцитируемости будет 2,0% (IEEE Transactions on Pattern Analysis and Machine Intelligence) и 9,9% (Journal of the European Ceramic Society). В частности, этот эффект приводит к тому, что исключение самоцитирования при вычислении импакт-факторов журналов (то есть 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, приведенная 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, приведенная 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 P_1 , которую цитирует статья P_2 , $P_2 \rightarrow P_1$. Это считается:

- прямым самоцитированием, если S — автор P_2 ;

- самоцитированием соавторов, если среди авторов P_2 есть хотя бы один из авторов P_1 .

Совсем упрощая, если статья Ловакова и Юдкевич цитирует работу, написанную Ловаковым и Писляковым, то для Ловакова это прямое самоцитирование, а для Пислякова — самоцитирование соавторов.

Соответственно, коэффициент самоцитируемости ученого может считаться или как доля «прямого самоцитирования» во всех

²³ В процессе подготовки переиздания «Руководства...» автор обнаружил серьезные изменения в показателях самоцитируемости российских журналов. Теперь (в JCR-2018) не найти изданий с самоцитируемостью 70% и выше, хотя ранее такие неоднократно встречались. Это обусловлено учетом с 2018 г. ссылок из базы ESCI при расчете библиометрических характеристик журналов JCR. Добавлен значительный корпус отечественной периодики, что привело к большей диверсификации получаемых журналами цитирований.

ссылках, которые получили его публикации, или как доля «самоцитирования соавторов» во всех ссылках, полученных его работами²⁴. 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: учтем только те цитирования, которые получены из «совсем чужих» работ — множество авторов которых никак не пересекается с множеством авторов цитируемой публикации. А все остальное посчитаем самоцитированием.

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 истории. Сложно представить себе автора, который бы не ссылался на свои предыдущие работы, это был бы либо ученый, 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 тот не отслеживает данный аспект.

3.3. 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-2018 она равна 3,49, а для математики — 0,92. Это означает, что половина журналов по клеточной биологии имеет импакт выше 3,49, в то время как половина журналов по математике получила импакт ниже 0,92. Более того, математический журнал с импакт-фактором 3,49 уверенно вошел бы в top-10 наиболее цитируемых журналов в своей рубрике (занял бы седьмое место из 314).

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:

²⁴ Заметим, 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, то есть не бывает «журналов-соавторов».

- Различная практика цитирования, сложившаяся в журналах различных научных областей; можно сказать, варьирующаяся в зависимости от дисциплины «плотность ссылочного поля». Например, известно, что в одной публикации по клеточной биологии в среднем 42 ссылки в пристатейном списке литературы, в то время как в работе по математике лишь 23 (JCR-2018). Поэтому «вероятность» получить ссылку у журнала по математике меньше²⁵. Это индивидуальные особенности коммуникации, укоренившиеся в различных науках и никак не относящиеся к их «качеству» 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. 3.2.2 и поговорим подробнее в разделе 3.8). Чем больше доля ссылок, ведущая 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. Например, в тех областях, где больше ссылок на книжные, а не журнальные источники; на труды конференций, диссертации, препринты, 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 публикаци-

онное окно, они также не увеличат импакт никакого журнала.

Эти дисциплинарные особенности не учитываются в абсолютных значениях импакт-индикаторов, что делает невозможным их 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.

3.3.1. Относительный импакт-фактор, относительная цитируемость

Начнем с относительного показателя для журнала. Самый простой из них — отношение импакт-фактора издания к среднему импакт-фактору той дисциплины, которая присвоена журналу. 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 так называемый «агрегированный импакт-фактор» (aggregate impact factor). Дисци-

²⁵ Разумеется, здесь мы используем приближенный подход, оставляя в стороне междисциплинарность и возможность получения ссылки «извне» дисциплины.

плина в целом (то есть совокупность статей, опубликованных во всех журналах, приписанных 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. Повторим, он отличается от арифметического среднего всех пожурнальных импакт-факторов в дисциплине. Посмотрите на приведенный пример.

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 описанный метод может помочь 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? Например, это могут быть статьи организации, в кото-

Пример

Поясним эту разницу в определении среднего на примере, поскольку она важна и встречается в разных библиометрических контекстах. В одной из самых маленьких тематических рубрик JCR-2018, «Андрология», всего 7 журналов. Их импакт-факторы равны 3,11; 2,86; 2,23; 2,11; 1,84; 1,72; 0,29. Их сумма равна 14,16, соответственно, среднее — **2,02**.

Но если взять дисциплину как целое, в ней во всех семи журналах за 2016–2017 гг. вышло 1333 статей, которые были процитированы в 2018 г. 3057 раз. Это дает агрегированный импакт-фактор $3057 : 1333 = 2,29$. Здесь мы получили 13 %-е различие в средних, для других рубрик JCR оно может быть и больше и меньше.

Агрегированный импакт-фактор, очевидно, также равен среднему числу ссылок, полученных одним журналом рубрики (усреднение по журналам), деленному на среднее число статей, вышедших в одном журнале (снова усреднение по журналам), при соблюдении обычных для импакт-фактора временных промежутков. Поэтому данный подход также называют *отношением средних*, в то время как первый, «арифметический», метод обозначают как *среднее отношение* (поскольку отдельный импакт-фактор сам по себе тоже отношение).

рой работает ученый, или более узко — статьи его лаборатории/факультета. Тогда мы узнаем, насколько эффективна деятельность данного ученого относительно его организации или на фоне коллег из 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 индикатора²⁶.

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 темати-

ческих областей, будет рассмотрен в п. 3.3.2.

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 вполне естественно, необходим более комплексный подход, который мы также рассмотрим в п. 3.3.2.

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

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

Представим себе теперь ситуацию, максимально приближенную к реальной: необходимо оценить научный уровень статей, опуб-

²⁶ При 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 ощутило увеличил число публикуемых статей в тот год, который предшествовал отчетному, он будет проигрывать журналу, не принимавшему таких редакторских решений (если предположить, что научный уровень всех статей в обоих журналах одинаков). Знаменатель первого журнала существенно вырос, а числитель не успел вобрать ссылки из-за малого истекшего времени. Эти недостатки преодолеваются при использовании более тщательного подхода, описанного в п. 3.3.2.

ликованных организацией широкого научного профиля, например университетом.

Сначала, как обычно, необходимо задать публикационное окно и окно цитирования, чтобы определить, какие именно публикации и на каком временном интервале мы оцениваем. 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 звучали в п. 3.3.1:

- определять «научный уровень» можно только на некотором фоне, в некотором контексте, *относительно чего-то*; то есть необходимо задать *референтную группу*;
- сравнивать научный уровень публикаций относительно друг друга можно, только если они *однотипны*, то есть *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: вклю-

чая публикации и самого анализируемого университета) — 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 изложения материала мы в дальнейшем будем считать референтной группой весь мир, то есть 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 уровня²⁷.

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 может получить к моменту измерения);
- тип публикации (как уже говорилось в п. 3.1.1, в среднем, например, научные обзоры цитируются чаще, чем оригинальные исследовательские статьи²⁸).

Для учета всех этих факторов и корректного определения относительной цити-

²⁷ Важно: 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.

²⁸ Этот факт далек от чистой теории. Например, документы типа Article, опубликованные в журнале Cell в 2014 г., получили за 2014–2018 гг. (из всех изданий Web of Science CC) в среднем по 128 ссылок. Тот же показатель для публикаций типа Review равен 203 — разница более чем в полтора раза. В редких случаях данная закономерность нарушается — так, для журнала Lancet аналогичные показатели равны 195 и 158, в пользу исследовательских статей. Так или иначе, учет типа документа не менее важен, чем коррекция на область науки, к которому он относится.

руемости всего публикационного потока организации/автора вводится показатель «ожидаемое число ссылок» (expected citation impact, 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 (то есть имеющей те же характеристики дисциплина/год выхода/тип документа), если меньше — уступает ему.

Относительная цитируемость одной публикации, таким образом, равна отношению реального числа полученных ею ссылок к их ожидаемому числу. Следуя терминологии InCites, обозначим этот показатель «Category Normalized Citation Impact» (CNCI)²⁹. Если статья опубликована в журнале, который приписан к двум или более тематическим рубрикам, ожидаемое число ссылок для каждой рубрики будет различным (ввиду разной средней цитируемости дисциплин). В этом случае CNCI считается как среднее арифметическое отношений числа полученных статьей ссылок к их ожидаемому числу в каждой из рубрик (среднее относительных цитируемостей в каждой дисциплине).

²⁹ В 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).

³⁰ Иногда используется альтернативный способ усреднения по всему массиву ссылок, полученных публикациями организации/автора. Сначала считается ожидаемое число ссылок сразу для всех статей (суммированием «ожиданий» по каждой), потом на эту сумму делится суммарное число реальных ссылок, полученное публикациями организации. Так определяется показатель «цитируемость, нормализованная по областям науки» в [Писляков, 2014]. Разница в этих методах усреднения абсолютно аналогична рассмотренному ранее в примере отличию «среднего отношений» и «отношения средних».

Наконец, для того чтобы посчитать CNCI совокупности всех публикаций организации/автора, берется среднее арифметическое значений CNCI для каждой отдельной публикации, написанной организацией/автором³⁰.

В методику расчета показателя CNCI 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 высокочитируемые статьи организации/автора. Даже для целых стран, как показали D. Aksnes и G. Sivertsen [Aksnes, Sivertsen, 2004], отдельные работы, получившие очень много ссылок, могут значительно повлиять на средние страновые показатели. С не меньшей осторожностью следует применять CNCI на коротких временных промежутках, например для статей 1–2-летней давности. Статья, удачно получившая на старте всего 3–4 ссылки, может в разы превзойти среднемировой уровень, и часто это будет недолгосрочный эффект. Кроме того, весь массив мировых публикаций получил еще недостаточно много ссылок к этому моменту, в результате чего показатели ожидаемой цитируемости могут быть малопредсказуемы и не давать надежного, достоверного базиса нормализации.

Наконец, как уже говорилось выше в сноске 27 на с. 198, CNCI, полученный относительно мирового уровня, не всегда дает возможность сравнивать исследуемые единицы между собой. Как рельефный пример — если у организации CNCI = 0,50, а у всей России CNCI = 0,80 (последняя цифра — реальное значение из InCites для статей 2014–2018 гг. выхода, с окном цитирования 2014 — по настоящее время), это не обязательно означает, что организация работает слабо относительно всей российской науки. Возможно, это институт, специализирующийся на информатике и вычислительных науках. Он публикует статьи лишь в два раза слабее мирового уровня, в то время как позиция России в Computer Science ощутимо ниже (0,44 от среднемирового уровня цитируемости). При этом значение CNCI = 0,80 для всей страны формируется благодаря более качественным публикациям по клинической медицине, строительству, психологии³¹.

3.3.3. Нормализация по журналам и комбинированный анализ относительной цитируемости

Подход, аналогичный представленному в п. 3.3.2, 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 исследуемой организации/автора

будет равно среднему числу цитирований, полученных теми публикациями данного журнала, которые вышли в том же году и относятся к тому же типу документа. Реальное 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 InCites, назовем Journal Normalized Citation Impact (JNCI)³². 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 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; Pislyakov, 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 исключе-

³¹ Данный пример иллюстративен, но охватывает лишь часть указанной проблемы: так, найдя значение CNCI для России с ограничением по вычислительным наукам, все равно некорректно сравнивать его с CNCI отдельной организации, специализирующейся на Computer Science: необходим комплексный подход с учетом и годов выхода статей, и типов документа. Предупредим, что имеющаяся в наборе индикаторов InCites метрика «Normalized Citation Impact (Country Adjusted)» не решает описанную задачу, она приводится для приблизительной (и скорее искусственной) коррекции показателей организаций из разных стран. Возник этот индикатор как ответ на требования рейтинговых агентств и на данный момент утратил актуальность.

³² Здесь вновь возможна иная нормализация, с подсчетом сразу всех ожидаемых ссылок, взятием суммы по всем статьям, а затем делением на полученный результат суммы всех полученных в реальности ссылок. Это соответствует метрике «Цитируемость, нормализованная по журналам» в [Писляков, 2014].

ниями (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. Все-таки при использовании последней в рамках одной области могут встречаться журналы, достаточно сильно отличающиеся по тематическим подкатегориям.

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 (CNCI), 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, следовательно, 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 все равно не достигает 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 CNCI, и JNCI. Суммируем выводы, которые можно сделать при их совмещенном анализе, в табл. 17. В качестве примера будем рассматривать публикации организации, хотя аналогичные результаты можно получить, исследуя другие научные единицы.

В дополнение к характеристикам, представленным в табл. 17, 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). В целом табл. 17 дает некоторое представление о том, какие журналы, насколько престижные издания, выбирают сотрудники — то, что называется «публикационной стратегией» ученого.

Исследование CNCI 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 покажет направление и темп эволюции научной деятельности ученых соответствующих вузов/институтов/лабораторий и т. д.

Анализ нормализованных показателей цитируемости для организации

	JNCI > 1	JNCI < 1
CNCI > 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*
CNCI < 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

*Осторожная формулировка «может быть связано» вызвана тем, что при нормализации, используемой сейчас в InCites, 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 CNCI и JNCI для организации. При использовании другого способа нормализации (например, [Писляков, 2014]) понятие «среднего уровня совокупности журналов относительно дисциплин» может быть формализовано более четко. Читатель может удивиться, отчего частному, на первый взгляд, вопросу способа нормализации («среднее отношений» vs. «отношение средних») уделяется столь много внимания. На самом деле этот вопрос — один из фундаментов в методиках расчета относительной цитируемости. В начале 2010-х гг. он стал предметом пространной и энергичной дискуссии ведущих мировых наукомеров в виде писем в редакцию Journal of Informetrics. См., например [Opthof, Leydesdorff, 2010; van Raan et al., 2010; Gingras, Larivière, 2011] и др.

3.3.4. 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, пятилет-

него импакт-фактора; или индекса влияния, который будет введен в разделе 3.6) 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, чет-

вертого 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, в рамках этого метода считаются 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 $n = 5$ «квинтиль» 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. Лидирующий журнал располагается в точке 1, самый последний — в точке 0, остальные равномерно размещаются между ними. Равномерно — несмотря на то, что по абсолютным значениям показателя какой-то журнал может быть гораздо ближе к своему верхнему соседу, чем к нижнему, или наоборот. Учитываются только ранги. Несложно вычислить, что при этом показатель «относительной позиции» журнала, занимающего n -е место в своем дисципли-

нарном рейтинге, равен $(N - n) / (N - 1)$. При кросс-дисциплинарном сравнении выигрывает тот журнал, у которого указанная «относительная позиция» выше. Это показано на рис. 36, 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): $\langle A1, B1 \rangle - B2 - A2 - B3 - \langle A3, B4 \rangle - B5 - A4 - B6 - \langle A5, B7 \rangle$.

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. п. 3.2.4, где ранговый подход, а именно использование медианы, предлагался для анализа импакт-факторов).

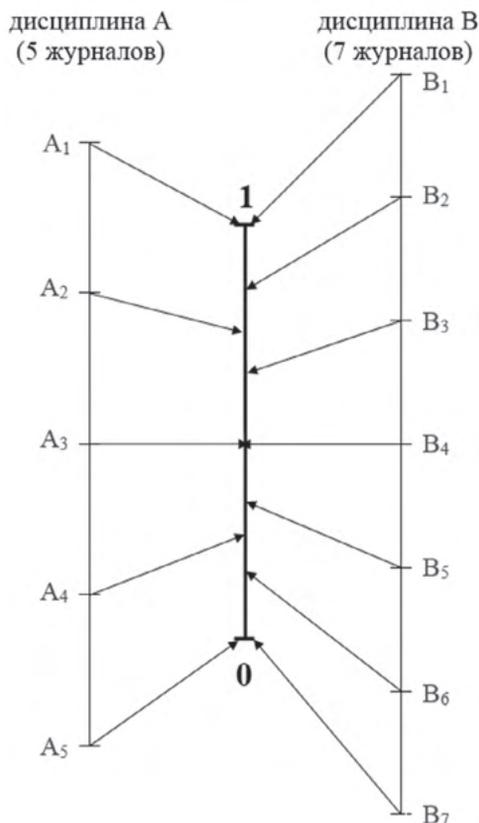


Рис. 36. Сравнение рангов журналов проекцированием на отрезок $[0; 1]$.
 $A_1 \dots A_5$ — журналы дисциплины А.
 $B_1 \dots B_7$ — журналы дисциплины В

В этом случае отказ от учета абсолютных значений индикаторов и исследование лишь мест в рейтингах может оказаться более эффективным.

3.4. Индикаторы «экстремальной» цитируемости

Еще один подход, коренным образом отличающийся от разобранных выше, заключается в том, чтобы не рассчитывать среднюю цитируемость всего массива ста-

тей, а заниматься «крайностями», наблюдаемыми в нем, — теми статьями, которые получили аномально малое или аномально большое 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 — так называемые «высокоцитируемые» статьи (*highly cited papers*), определение которых дается в п. 3.4.1.

Отличительной чертой «экстремального» 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.

3.4.1. 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 исследовательских 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 рассчитываются пороговые значения числа цитирований, которое должна получить к настоящему моменту статья, чтобы попасть в 1% высокоцитируемых. Как и ранее в п. 3.3.2, 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 март 2020 г., высокоцитируемой статьей 2010 г. выхода по дисциплине «Микробиология» была лишь та, которая получила за прошедшие 10 лет не менее 230 цитирований, в то время как статье по математике достаточно было получить 81 ссылку. Аналогичные пороги для публикаций 2015 г. выхода по этим дисциплинам составляли 100 и 35 ссылок соответственно³³.

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 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 некоторый 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

³³Заметим, что, в отличие от методов, разобранных в п. 3.3.2, нормализация в ESI не затрагивает тип документа. При этом в ESI обрабатываются только документы типа Article и Review.

коллектива (каждая сотая статья) достигает категории высокоцитируемости, это говорит о соответствии деятельности ученых мировому уровню³⁴.

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.

3.4.2. 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 коллектива³⁵.

Теперь предположим, что некоторая статья не получила на текущий момент 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 2010 г. и не процитированных до 2014 г. включительно; опубликованных в 2011 г. и не процитированных до 2015 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, вышедших в течение 2010–2013 гг., увеличивалась или уменьшалась (хотя для статей каждого года выхода берется свой временной интервал, на котором проверяется отсутствие ссылок). Для сравнения различных организаций или авторов между собой необходимо оставаться в рамках одного дисциплинарного направления, поскольку среднемировой коэффициент нецитируемости, как и остальные ненормализованные показатели, сильно зависят 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

³⁴ Строго говоря, ситуация несколько сложнее. Высокоцитируемые статьи часто пишутся в больших соавторских коллективах, и для корректного соотнесения с мировым уровнем следует использовать дробный счет (fractional counting, см. п. 3.1.1). Так, в работе [Pislyakov, Shukshina, 2014] найдено, что при подсчете whole counting российские ученые опубликовали 927 высокоцитируемых статей, но если вести дробный счет, это число уменьшится втрое. 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.

³⁵ Зеркальный взгляд на это явление реализован в InCites, 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 InCites используется не нормализованный по времени показатель (то есть не такой, как вводится далее в данном разделе).

и SSCI), составил около 40%. Это означает, что для получения осмысленного показателя массиву публикаций достаточно содержать сотни или даже десятки статей.

3.5. Профили цитируемости

Распространением и обобщением подходов «экстремальной цитируемости» является исследование долей статей организации, или автора, или любой другой научной единицы, попадающих в каждый процентиль по цитируемости среди себе подобных статей. Возьмем одну статью и все множество «подобных» ей — как обычно, это обозначает статьи той же тематики, того же года выхода и того же типа документа. Упорядочим весь массив в порядке *возрастания* числа полученных статьями ссылок, после чего разделим этот упорядоченный список, например, на 10 равных по числу статей интервалов (часто используется именно такое деление, каждый интервал называется децилем). Первые 10% публикаций будут нецитируемыми и наименее цитируемыми, последние 10% самыми цитируемыми (а десятая часть этих статей будут высокоцитируемыми по определению раздела 3.4.1, так как они попадут в интервал 99–100%). Теперь можно посмотреть, в какой дециль попала взятая нами отдельная статья. Скажем, если всего «подобных» ей статей вышло 10 000, в списке по возрастанию цитируемости статья находится на 3451-м месте, значит, она попадает в дециль 30–40%³⁶.

³⁶ Существуют разные подходы для определения дециля статьи, если опубликовавший ее журнал (а следовательно, сама статья) относится к нескольким тематическим категориям. Децили по разным категориям могут не совпадать. Часто в таком случае берут лучший показатель — наиболее высокий дециль.

³⁷ Обратим внимание читателя, что профилирование цитируемости сейчас является одним из наиболее грамотных подходов, имеющих высокий авторитет в наукометрическом сообществе (см., например, отчет «Profiles, not metrics» — Adams et al., 2019). Однако ввиду того, что результатом является не «одно понятное число», а график профиля, требующий некоторых усилий для содержательного анализа, этому методу, к сожалению, редко дают ход в официальных отчетах, формах грантовых заявок и других административных документах. Данный раздел написан специально для второго издания «Руководства», чтобы рассказать аудитории об актуальных методах оценки научной деятельности, находящихся на переднем крае современной наукометрии.

Далее — и в этом суть метода — можно взять все публикации автора/организации/страны/региона (конечно, лучше ограничиться вышедшими за какой-то определенный временной промежуток), узнать дециль, в который попадает каждая статья и получить суммарную раскладку по исследуемому объекту: сколько его статей попало в дециль 0–10%, сколько в 10–20%, ..., сколько в наиболее цитируемый дециль 90–100%. Заметьте, для каждой статьи дециль будет определяться по своему множеству «ей подобных» (с теми же характеристиками тематики, года выхода и типа), однако результаты — полученные децили — можно, и это будет вполне корректно, объединять по всему массиву публикаций.

Итогом исследования является так называемый профиль цитируемости автора/организации — гистограмма (или, в сглаженной форме, график) распределения долей опубликованных автором/организацией статей по децилям цитируемости. Построение этой диаграммы позволяет наглядно увидеть, как часто организация/автор публикуют сильные, средние, слабые статьи. Это именно полная картина библиометрической «эффективности», «результативности» публикаций ученого³⁷.

Еще раз подчеркнем, что речь идет о профилях нормализованной по рангу цитируемости, когда каждая статья аккуратно рассматривается относительно себе подоб-

Пример

Сначала возьмем воображаемый простой пример, чтобы показать механику расчета данных для диаграммы профиля. Пусть ученый написал в определенном году две исследовательские статьи (Article) — одну по математике, одну по статистике. На данный момент они получили 3 и 7 ссылок соответственно.

При этом в математике всего во всем мире в том же году вышло 11 статей (включая статью данного ученого), с цитируемостью на текущий момент 0, 0, 1, 2, **3**, 5, 6, 6, 7, 9, 9. По статистике — 12 (пример, напомним, полностью воображаемый), с цитируемостью 0, 0, 0, 1, 1, 5, 5, 6, **7**, 9, 9, 15. Первая статья исследуемого ученого, занимая 5-е из 11 мест по цитируемости, входит в ранговый интервал 40–50%. Вторая, занимая 9-е место из 12, входит в интервал 70–80%. Гистограмма профиля цитируемости для данного автора будет состоять из столбца 50% (поскольку одна статья — это половина всех публикаций автора, а график, напомним, это доли, проценты от всех публикаций ученого) в дециле 40–50% и столбца 50% в дециле 70–80%.

ного окружения. Это значит, что допустимы 1) сравнение профилей между разными дисциплинами; 2) объединение статей всех авторов лаборатории/факультета/организации в единый профиль; 3) рассмотрение публикационного потока за несколько лет (напомним, нормализация включает в себя и корректировку на год выхода).

Посмотрите приведенный пример. Реальные данные, конечно, всегда гораздо масштабнее, однако расчеты дециля, в который попадает каждая статья по нормализованной цитируемости, легко осуществить при помощи базы данных InCites. Если по ней провести поиск всех работ автора или организации, то для каждой публикации будет указан индикатор «percentile in subject area». Это в своем роде обратный показатель — он равен проценту статей, которые превосходят данную публикацию по цити-

руемости. То есть значение «0» — это самая цитируемая статья среди себе подобных («никто не сильнее меня»), 100 — самая низкоцитируемая («я слабее всех»). Это не сложно перевести в децили: если percentile in subject area равен 0–0,099, это соответствует самому высокому децилю 90–100%, если равен 0,1–0,199, то децилю 80–90% и т. д.³⁸. Пользуясь таким методом, мы легко построим уже реальный профиль цитируемости (рис. 37). Это публикации Национального исследовательского университета «Высшая школа экономики» за 2015–2017 гг. Взятые только типы документов Article и Review и только «старшие» журнальные индексы, ESCI исключен.

График показывает, что чуть менее 10% статей вуза попали в лучшие 10% по цитируемости в своей области (дециль 90–100), 8% оказались в следующих 80–90% по ци-

³⁸ К сожалению, эта «противонаправленность» двух одинаковых по сути показателей присуща ранговым исследованиям. В ее основе лежит неизбежное противоречие между тем, что самые хорошие публикации/журналы мы хотим называть *первыми* («первый квартиль», «первое место» и др.), но при этом они получают *максимальное* число ссылок или какой-то другой меры престижа. «Первое» (место) по обозначающему его числу меньше «второго», но показатель, соответствующий ему, выше, больше. Практический совет: этот нюанс может иногда серьезно вводить в заблуждение, поэтому при работе с разными базами данных, использующими квартили, децили, проценти и пр., всегда необходимо выяснить, *откуда исчисляется порядковый номер*. Например, является ли принадлежность «первому квартилю» знаком лучших показателей (как было у нас с журналами в п. 3.3.4) или, наоборот, это «хвост» распределения, самые низкие значения. Все это может зависеть от обозначений, принятых в конкретном библиометрическом инструменте.

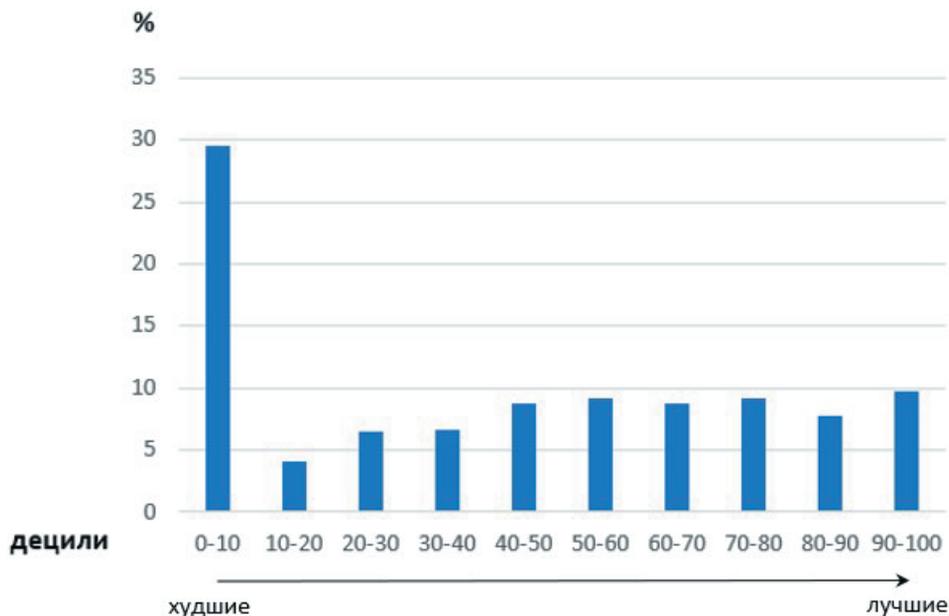


Рис. 37. Профиль цитируемости НИУ ВШЭ (2015–2017). Типы документов: Article, Review

тируемости и т. д. Наконец, около трети работ попали в наименее цитируемый дециль. Заметьте, что мы намеренно взяли достаточно удаленный интервал 2015–2017 гг., поскольку внятное ранжирование статей по цитируемости возможно лишь по истечении определенного времени.

3.6. «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 приводит 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 высокочитируемого журнала, например журнала Nature, «ценится» выше, чем ссылка из малоизвестного регионального издания. Аналогичная идея лежит в основе расчета ранга «авторитетности» веб-страниц некоторыми поисковыми машинами, например Google (так называемый PageRank).

Взвешенный журнальный индикатор, получивший название «собственный фактор» (Eigenfactor), был предложен в 2007 г. специалистами из лаборатории Карла Бергстрема (Университет Джорджа Ва-

шингтона). Разработчики данного показателя иллюстрировали его с помощью достаточно долгого (бесконечного) 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 наличие многих «путей» (ссылок) в этот журнал, но также надо, чтобы приходили эти ссылки из часто посещаемых читателем журналов.

Еще один механизм, которым иллюстрируют взвешенные показатели, — это «раздача престижа» журналом в результате итеративной процедуры. Сначала каждому журналу назначается одинаковый начальный «престиж», равный единице. На первом этапе итерации каждый журнал поровну «делит» свой единичный престиж между всеми ссылками. Он «посылает» ссылки другим журналам, передавая таким образом долю своего престижа. Чем меньше в журнале сделано ссылок, тем больше «вес» каждой из них³⁹. Таким образом рассчитывается «престиж», полученный каждым журналом из других журналов. Это значение престижа, полученного в результате первой итера-

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

Согласно строгому математическому определению, для вычисления значения собственного фактора рассматривается марковский процесс, 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* (отчетного) года на публикации, вышедшие за *пять предыдущих лет*. Таким образом, публикационное окно и окно цитирования совпадают с установленными для пятилетнего импакт-фактора. Особенностью этих показателей является то, что ссылки журнала на свои же статьи не учитываются, — в отличие от импакт-фактора, самоцитирование из расчетов исключено.

Использование взвешенных индикаторов в совокупности с традиционным импакт-фактором (двух- или пятилетним) позво-

³⁹ Точно 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.

ляет 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, который подозревался во «взаимном договорном цитировании», в JCR-2018 занял 18-е место из 363-х по импакт-фактору в дисциплине «Экономика». Если исключить самоцитирования, журнал останется в лидерах — 20-е место. И только если обратить внимание на показатель Afrticle influence, окажется, что он на 210-м (!) месте.

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 ниже, чем в тех областях науки, где ссылок в расчете на статью меньше и «ссылочное поле» менее плотное. Кроме того, пятилетнее пуб-

ликационное окно также способствует более равносному учету цитирований в различных научных дисциплинах.

3.7. 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 (то есть 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 3.4). Это имеет свой смысл, поскольку усреднение зачастую не дает полной картины об исследуемом множестве статей и не позволяет производить корректные сравнения 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 сначала 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.

3.7.1. Индекс Хирша

Согласно определению, индекс Хирша массива публикаций ученого равен h , если есть h статей из этого массива, каждая

Пример

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 раз.

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.

Графическая иллюстрация методики расчета приведена на рис. 38. Каждая точка графика соответствует одной публикации ученого. Публикации упорядочены по цитируемости: от тех, которые получили больше всего ссылок, к наименее цитируемым статьям. По оси абсцисс на рис. 38 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 написал 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 получила 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 на любой набор статей, для каждой из которых известно число полученных ссылок.

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, только расти или оставаться неизменным⁴⁰, — и ученый может уже много лет на-

⁴⁰ Строго 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) —

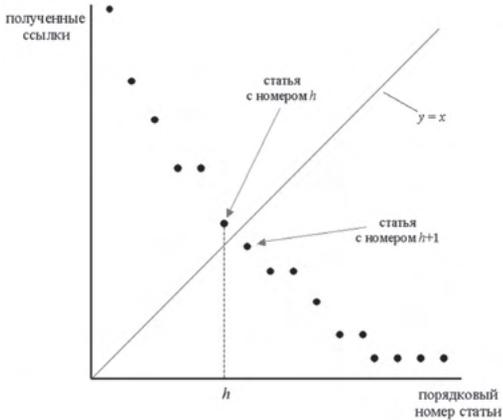


Рис. 38. Индекс Хирша (h -индекс)

ходить 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 хорошей «разрешающей способностью» (были предложены способы обобщить h -индекс на дроб-

ные 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. 38 получила бы в 10 раз больше ссылок (а цитируемость остальных публикаций осталась прежней), на значении h -индекса это бы никак 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 «научного 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. В целом в трудах ведущих библиометров в последнее время интерес и внимание к индексу Хирша заметно упали, его все реже рассматривают и изучают

из-за найденных серьезных ошибок или обнаружения плагиата и т. д. На это обстоятельство обращал внимание В. Глэнцел (конференция ISSI 2009, устная реплика).

как некий серьезный, несущий глубокий смысл показатель. В этом контексте актуальность раздела 3.7 настоящей главы снизилась по сравнению с первым изданием. Тем не менее давайте рассмотрим две известные попытки модифицировать, в чем-то «улучшить» h -индекс, возникшие в первые годы после его изобретения Х. Хиршем.

3.7.2. 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, N.J. Van Eck и L. Waltman [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 менее αh_α цитирований, а каждая из остальных — не более αh_α цитирований.

Графически смысл h_α -индекса отражен на рис. 39. 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. 39 показаны примеры для $\alpha=2$ и $\alpha=1/2$.

Основной смысл нового индикатора в следующем. Как мы говорили, индекс Хирша одновременно оценивает число публикаций и их цитируемость, делая это в каком-то

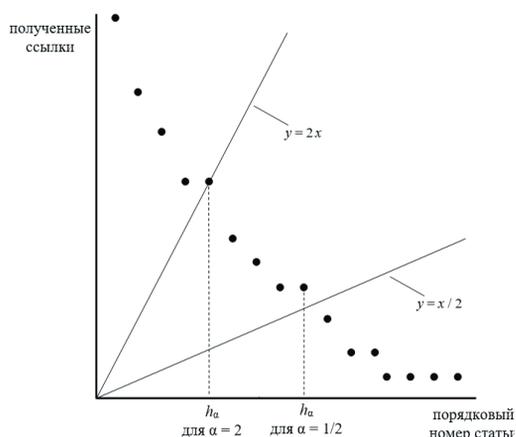


Рис. 39. h_α -индекс. Пример для $\alpha = 2$ и $\alpha = 1/2$

Пример

Вспомним ситуацию: один автор написал 5 статей, каждая из которых получила по 100 ссылок, второй 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). В первом случае важнее наличие высокоцитируемых статей, во втором — высокая продуктивность ученого⁴¹.

⁴¹Если рассмотреть предельные случаи, то для любого автора максимальное значение α , при котором его h_α -индекс все еще больше нуля (то есть равен 1), совпадает с числом цитирований, которое получила его самая высокоцитируемая статья. Число статей при этом не играет никакой роли. Наоборот, при уменьшении 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»).

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 отклонением от этой «равномерности». 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.

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, то есть на числе публикаций, или делая основной акцент на качественном аспекте, на цитируемости изучаемых работ.

3.7.3. 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 рассматривается совокупная цити-

руемость 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.

Можно переформулировать определение g -индекса так: g -индекс ученого равен такому максимальному g , что у ученого есть g статей со *средней цитируемостью* 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 здесь примере самая цитируемая статья со временем получит еще три ссылки (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.

Пример

Вернемся к примеру из п. 3.7.1, 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, $g \geq h$.

3.8. 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 ранее значения медианы, половина — на более свежую литературу.

Время полужизни ссылок (см. пример) зависит от научной области, к которой относится журнал. Оно и характеризует эту область с точки зрения темпов устаревания производимого в ней научного знания. В JCR-2018 минимальное время полужизни сделанных ссылок у журналов дисциплин «Электрохимия» и «Нанонауки и нанотехнологии» (5,7 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-2018 в общей

Пример

На рис. 40 показано хронологическое распределение цитирований, сделанных в статьях журнала *Annual Review of Biochemistry*, вышедших в 2018 г. В этих статьях было 43 ссылки на публикации всех журналов WoS, вышедших в 2018 г., 409 ссылок на источники, вышедшие в 2017 г., и т. д. В общей сложности статьи 2018 г. этого журнала содержали 5621 ссылку, из них половина пришлась на документы, вышедшие в 2011–2018 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 — термин, родственному используемому в биологических науках, атомной физике и легендарной компьютерной игре). Для данного журнала в 2018 г. время полужизни сделанных ссылок равно 7,9 лет (затененной областью на рисунке отмечены ссылки на более поздние источники⁴²).

⁴² Из-за того, что JCR лишь схематически показывает этот индикатор на диаграмме, «тень» приходит в центр столбца гистограммы, соответствующего 2012 г. Это не совсем точно, правильно отмечать границу на правом краю столбца 2011 г. (то есть правее, чем сейчас). Точное значение citing half-life публикуется в численном виде в JCR. На визуализации присутствуют данные только за 10 последних лет, в базе данных значения старше 10 лет объединяются в одно число.

Citing Journal Graph 2018

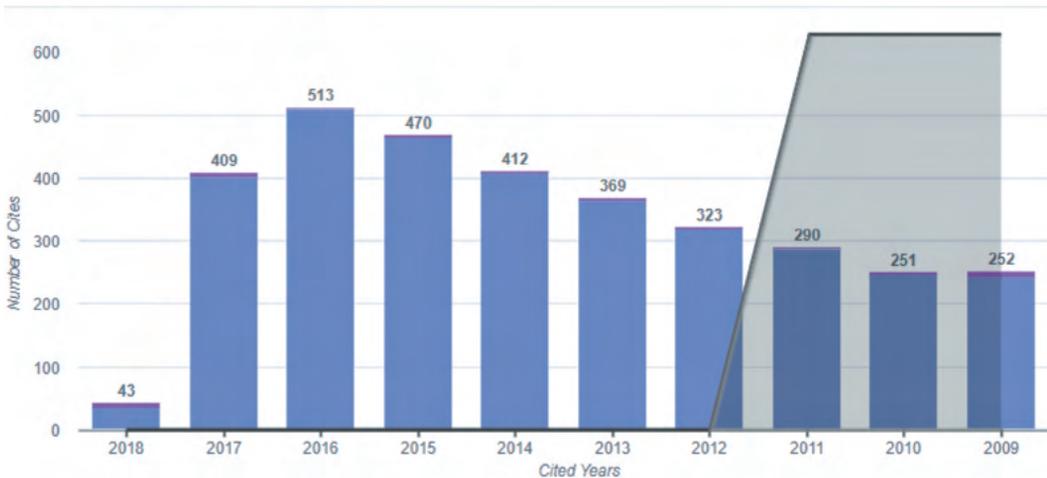


Рис. 40. Хронологическое распределение ссылок в статьях журнала Annual Review of Biochemistry, вышедших в 2018 г. Скриншот интерфейса JCR (Clarivate)

сложности 53, среди них, в частности, культурология, 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 вводится не только для сделанных ссылок, но и, 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 востребованы прежде всего свежие, недавние публикации издания. Если велика — полезными оказываются и достаточно старые материалы, когда-то публиковавшиеся в журнале. Время полужизни полученных ссылок для целых научных дисциплин варьируется в JCR-2018 от трех лет («Науки и технологии,

сохраняющие окружающую среду») до более 10 лет (49 дисциплин). Безусловно, *cited half-life* связан с *citing half-life*: напрямую они не зависят друг от друга, однако оба показателя отражают скорость эволюции знания в соответствующей научной области. Поэтому, например, среди 49 дисциплин со временем полужизни полученных ссылок более 10 лет в 40 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 ученых, научных коллективов, организаций и целых стран. Были также изучены разнообразные журналь-

ные показатели. Заметим, что это далеко не исчерпывающий список индикаторов, существующих в библиометрии, и даже не исчерпывающий перечень индикаторов, доступных в инструментах компании Clarivate. Как не имеющие прямого отношения к цитируемости и оценке эффективности научной деятельности, нами были оставлены в стороне вопросы, касающиеся измерения международного сотрудничества в публикациях, характеристик сетей соавторства, индексов научной специализации, коэффициентов 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 настоящей главе 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 многосторон-

не 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 финансирования, 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. : НПБК; 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 // Научно-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.

Adams J., McVeigh M., Pendlebury D., Szomszor M. Profiles, not metrics. Clarivate Analytics, 2019. — <https://>

clarivate.com/webofsciencegroup/wp-content/uploads/sites/2/dlm_uploads/2019/07/WOS_ISI_Report_ProfilesNotMetrics_008.pdf

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

Callaway E. Publishing elite turns against impact factor // *Nature*. — 2016. — Vol. 535(7611). — P. 210–211. DOI: 10.1038/nature.2016.20224.

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

Gingras Y., Larivière V. There are neither “king” nor “crown” in scientometrics: Comments on a supposed “alternative” method of normalization // *Journal of Informetrics*. — 2011. Vol. 5, Is. 1. — P. 226–227.

Grančay M., Vveinhardt J., Šumilo Ě. Publish or perish: How Central and Eastern European economists have dealt with

the ever-increasing academic publishing requirements 2000–2015 // *Scientometrics*. – 2017. – Vol. 111, Is. 3. – P. 1813–1837. DOI: 10.1007/s11192-017-2332-z.

Hirsch J. E. An index to quantify an individual's scientific research output // *Proceedings of the National Academy of Sciences*. – 2005. – Vol. 102, Is. 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, Is. 8. – P. 997–1004.

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

Ophthof T., Leydesdorff L. Caveats for the journal and field normalizations in the CWTS (“Leiden”) evaluations of research performance // *Journal of Informetrics*. – 2010. Vol. 4, Is. 3. – P. 439–440.

Pislyakov V., Dyachenko E. Citation expectations: are they realized? Study of the Matthew index for Russian papers

published abroad // *Scientometrics*. – 2010. Vol. 83, Is. 3. – P. 739–749.

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

Seglen P. O. Why the impact factor of journals should not be used for evaluating research // *British Medical Journal*. – 1997. – Vol. 314(7079). – P. 498–502. DOI: 10.1136/bmj.314.7079.497.

Tregoning J. How will you judge me if not by impact factor? // *Nature*. – 2018. – Vol. 558(7710). – P. 345.

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

Van Raan A. F. J., van Leeuwen T. N., Visser M. S., van Eck N. J., Waltman L. Rivals for the crown: Reply to Ophthof and Leydesdorff // *Journal of Informetrics*. – 2010. – Vol. 4, Is. 3. – P. 431–435.

Yuret T. Is it easier to publish in journals that have low impact factors? // *Applied Economics Letters*. 2016. – Vol. 23, Is. 11. – P. 801–803. DOI 10.1080/13504851.2015.1109034.

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.

Иванов В. В., Маркусова В. А., Миндели Л. Э. Государственные инвестиции и публикационная активность вузов: библиометрический анализ // *Вестн. РАН*. – 2016. Т. 86, № 7. – С. 611–619.

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.

Clarivate. InCites Indicators Handbook. 2018. – <https://incites.help.clarivate.com/Content/Resources/Docs/indicators-handbook-june-2018.pdf>.

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, No. 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, No. 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, Is. 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, Is. 3. – P. 379–391.

Guskov A. E., Kosyakov D. V., Selivanova I. V. Boosting research productivity in top Russian universities: the circumstances of breakthrough // *Scientometrics*. – 2018. – Vol. 117, Is. 2. – P. 1053–1080. DOI: 10.1007/s11192-018-2890-8.

Handbook of Quantitative Science and Technology Research / H. F. Moed, W. Glänzel, U. Schmoch (eds.). – 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, Is. 3. – P. 265–277.

Moed H. F., Markusova V., Akoev M. Trends in Russian research output indexed in Scopus and Web of Science // *Scientometrics*. – 2018. – Vol. 116, Is. 2. – P. 1153–1180. DOI: 10.1007/s11192-018-2769-8.

Pislyakov V. Comparing two «thermometers»: Impact factors of 20 leading economic journals according to Journal Citation Reports and Scopus // *Scientometrics*. – 2009. – Vol. 79, Is. 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, Is. 11. – P. 2321–2330. DOI: 10.1002/asi.23093.

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

Springer Handbook of Science and Technology Indicators / H. F. Moed, W. Glänzel, U. Schmoch, M. Thelwall (eds.). – Switzerland: Springer Nature, 2019.