
**МАТЕМАТИЧЕСКИЕ МЕТОДЫ
И МОДЕЛИРОВАНИЕ В ПРИБОРОСТРОЕНИИ**

УДК 025.04.036:681.2

© Д. Б. Архипов, А. Л. Буляница, А. П. Щербаков, 2019

**АНАЛИТИЧЕСКОЕ ПРИБОРОСТРОЕНИЕ В ЖУРНАЛАХ
NATURE И SCIENCE ЗА 2001–2017 гг.
ВЕБОМЕТРИЧЕСКИЙ АНАЛИЗ**

Проведен вебметрический анализ журналов Nature и Science за 2001–2017 гг. По каждому году выбраны 10 наиболее цитируемых статей. Число ссылок определялось по Web of Science и Google Scholar. Было установлено, что из 170 статей не менее 85 имеют прямое отношение к аналитическому приборостроению. Формальным критерием выбора было наличие в статье упоминания какого-либо аналитического прибора (например, секвенатора). Также была проанализирована динамика цитирования статей и предложена легко интерпретируемая математическая модель процесса цитирования.

Кл. сл.: вебметрический анализ, высокоцитируемые публикации, аналитическое приборостроение, динамика цитирования, математическая модель, число ссылок

ВВЕДЕНИЕ

Журналы Nature (Лондон), выпускаемый с 1869 г. и Science (Вашингтон), выпускаемый с 1880 г., входят в десятку наиболее цитируемых журналов мира. В отличие, например, от Physical Reviews и Cell эти журналы доступны для неспециалистов. Классический наукометрический анализ малоэффективен для оценки распределения статей по областям науки, поскольку доли статей, относящихся к различным областям, таким как астрономия, физика, химия, науки о Земле, биология и гуманитарные науки сопоставимы. При этом число ссылок на одни статьи измеряется десятками, а на другие — тысячами.

В XXI в. Nature и Science публиковали 800–850 статей в год. Для определения числа ссылок использовались международные библиометрические базы Web of Science и Google Scholar (методика описана в статье [1]).

Основные наблюдаемые закономерности.

1. В среднем цитирование в Google на 50% выше, чем в Web of Science, но бывают исключения.

2. По сравнению с 1965 г. число статей в Nature упало в 4 раза. Это связано с тем, что кроме этого журнала под названием Nature в настоящее время выходит около 30 журналов, например NaturePhotonics или NatureMicrobiology.

Проблема старения научной литературы заслуживает отдельного исследования. В настоящее время наиболее цитируемой статьей (214 524 ссылки по Google на 15.11.2018 г.) является коло-

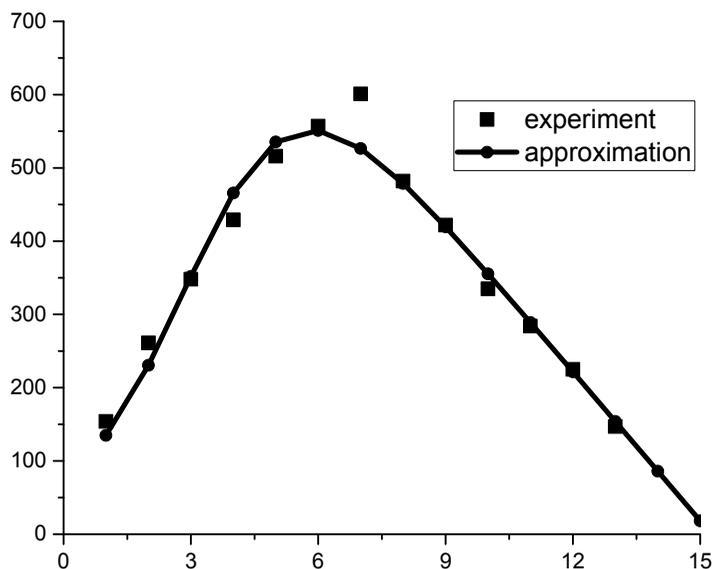
риметрический метод количественного определения белков в растворе, предложенный Лоури в 1952 г. [2]. Вместе с тем большинство статей в Nature и Science за 1977 г. цитируются скромно по сравнению с современными. 24.02.1977 г. Сенгер опубликовал в Nature две статьи по применению геле-электрофореза для секвенирования ДНК. Одна из них имеет 1955 ссылок, другая — 54. В декабре того же года он опубликовал в Proceedings of the National Academy of Sciences методическую статью по "методу Сенгера" [3], хотя в те годы английские ученые редко публиковались в американских журналах. Эта статья на 15.11.2018 имела 73 301 ссылку по Google и занимала четвертое место по цитированию.

ЗАДАЧИ. МЕТОДИКА РЕШЕНИЙ

В статье использованы два способа определения числа ссылок. В первом: к числу ссылок по Web of Science прибавляется число ссылок по Google Scholar. Во втором способе число ссылок в Web of Science умножается на 1.5. Оба этих способа имеют аргументацию в свою пользу.

При анализе динамики числа публикаций (распределения числа цитирований по годам после выхода статьи) и при оценке связи между цитируемостью наиболее цитируемой и десятой по цитируемости статьями по каждому году можно использовать условные (нормированные) показатели цитирования.

При решении первой задачи ключевой является зависимость от времени, во втором случае исполь-



Динамика ежегодного числа цитирований статьи [4].

По оси абсцисс — годы (число лет) после публикации, по оси ординат — число ссылок

зуется корреляционный анализ. Как известно, коэффициент корреляции инвариантен к изменению масштаба и линейному сдвигу шкалы отсчета.

Ранее для других групп публикаций авторами была предложена математическая модель [1] зависимости цитирования (числа ссылок) от времени. Формально эта зависимость представляет собой произведение убывающей линейной функции (учет морального старения информации) и логистической функции первого порядка, позволяющей описывать процесс распространения информации (аналогично динамике развития спроса на товары). Задачей авторов было оценить правомерность применения этой модели к специальной группе научных статей — статей, имеющих высокое цитирование.

При решении второй задачи на основе вычисленного коэффициента корреляции проверяется следующая аналогия: если число цитирований статьи можно рассматривать как случайную величину, то ранжирование статей по числу цитирований представляет собой процесс упорядочивания выборки (построение вариационного ряда). Таким образом, вычисляется коэффициент корреляции между первой и 10-й порядковыми статистиками. Всего элементов упорядоченной выборки порядка 800–850. Из теории известно, что близкие друг к другу порядковые статистики должны быть положительно коррелированы.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

В 2001–2017 гг. Nature и Science суммарно опубликовали в общей сложности около 28 тыс.

научных статей. В перечень для рассмотрения попали по 10 наиболее цитируемых статей за каждый год плюс самая цитируемая статья за декабрь 2017 г., что составляет 0.6 %.

Из рассматриваемого набора 170 высокоцитируемых статей для анализа динамики цитирования и проверки адекватности ранее описанной математической модели [1] выбрана публикация [4] 2005 г. Во-первых, эта статья относится к средним по цитированию работам из группы рассмотренных, во-вторых, период цитирования 2006–2018 гг. позволяет оценить динамику ежегодного числа ссылок и оценить параметры модели с достаточной точностью.

Зависимость числа ссылок на публикацию и аппроксимирующая зависимость представлены на рисунке.

Математически указанная зависимость выражается формулой:

$$X[N] = (1 - N/15.27) \cdot 1037.35 / (1 + 13.16 \exp(-0.754N)).$$

Здесь N — число лет после публикации, $X[N]$ — число цитирований в N -й год после публикации. Коэффициент детерминации зависимости превышает 0.93.

Значение характерного временного интервала в модели — 15.27 г. — характеризует время практически полной потери интереса (актуальности) к публикации. Ранее авторами были рассмотрены аналогичные модели активности цитирования на примере журналов Американского химического общества (АХО) для трех групп публикаций: лидеров цитирования, статей со средним цитированием и аутсайдеров. Соответствующие величины временного показателя равны 73.35, 30.20 и 21.60 г.

Табл. 1. Распределение уровня цитируемости по годам за период 2001–2017 гг. для двух выбранных публикаций (наиболее цитируемой и 10-й по цитируемости)

Год	Количество ссылок		Год	Количество ссылок	
	Наиболее цитируемая статья	10-я по цитируемости		Наиболее цитируемая статья	10-я по цитируемости
2001	42662	10930	2010	12060	5431
2002	24505	13595	2011	11930	4860
2003	15824	8420	2012	16320	6125
2004	89919	8003	2013	13617	6580
2005	35850	9450	2014	8530	4439
2006	23210	11445	2015	7443	3581
2007	18341	8172	2016	6405	1744
2008	27356	7602	2017	3139	799
2009	20772	7191	—	—	—

Табл. 2. Распределение числа статей по тематике из перечня высокоцитируемых статей для различных временных периодов

Область знаний	Временной период, годы		
	2001–2006	2007–2012	2013–2018
ДНК и РНК	20	28	18
Солнечные батареи	0	2	15
Физика	16	19	3
Молекулярная биология и медицина	10	6	4
Химия	4	1	4
Экология	5	1	2
Всего:	60	60	50

Т.е. в этом исследовании интерес к тестовой публикации [4] убывает (до нулевого цитирования) существенно быстрее, чем для статей АХО всех групп цитирования.

В табл. 1 приведены данные по числу ссылок на 1-ю и 10-ю по цитированию статьи в каждый из рассматриваемых годов публикации (с 2001 до 2017 г.)

Вычисление коэффициента корреляции производилось по двум методикам: выборочный коэффициент корреляции Пирсона и ранговый коэф-

фициент корреляции Спирмена. Оценка выборочного коэффициента корреляции равна +0.478, рангового — (+0.848).

Столь сильное расхождение объясняется неробастностью (малой устойчивостью к выбросам) выборочного коэффициента корреляции. Среди элементов выборки по максимальному цитированию есть аномально высокое значение — 89 919. Если его исключить, то оценки коэффициентов корреляции уточнятся и сблизятся по величине: по Пирсону +0.800, по Спирмену +0.900. И та, и дру-

гая оценки свидетельствуют о наличии значимо большой положительной корреляции, что соответствует теории вероятностей и статистики, а именно своим свойствам порядковых статистик.

Другой задачей было исследование распределения числа (доли) высокоцитируемых статей по различным областям знаний и его динамики в зависимости от периода публикаций. Основные результаты иллюстрируются данными табл. 2.

Из данных табл. 2 видно, что наибольшее число высокоцитируемых статей приходится на анализ нуклеиновых кислот, невозможный без определения последовательности нуклеотидов. Во всех способах секвенса необходима полимеразная цепная реакция (ПЦР). Первая статья по ПЦР была опубликована в журнале Science в ноябре 1985 г. Затем были изобретены ПЦР в реальном времени, цифровая ПЦР и цифровая ПЦР в капле [5]. Практически во всех приборах ПЦР используется флуориметрический детектор. До 2005 г. все секвенаторы использовали гель-электрофорез, вначале плоскостной, затем в капиллярах, а с середины 1990-х гг. — в чипах. В сентябре 2005 г. журнал Nature опубликовал статью по усовершенствованному пиросеквенсу. В этом приборе вместо электрофореза используются биоспецифические методы разделения, а вместо флуоресцентного детектирования — биOLUMИнесцентное. В пионерской работе по определению полного генома человека (Джеймса Уотсона) был использован именно пиросеквенатор [6].

Три основных наблюдаемых тенденции.

А. Многократное возрастание числа статей, относящихся к солнечным батареям.

Б. Число статей, касающихся других разделов физики, уменьшается. Однако суммарно число статей по указанным обеим группам примерно стабильно составляет около трети всех высокоцитируемых публикаций.

В. Устойчивое убывание числа статей, относящихся к молекулярной биологии. При этом динамика публикаций по РНК и ДНК, химии и экологии носит колебательный характер.

Для изучения работы клетки необходимо знать не только последовательность нуклеотидов, но и последовательность аминокислот в белках. Для решения этой задачи используют капиллярную жидкостную хроматографию в сочетании с тандемной масс-спектрометрией (МС) и дополнительно МС с матрично-активированной лазерной десорбцией/ионизацией (МАЛДИ). Для определения вторичной структуры белков в настоящее время используют криоэлектронную микроскопию.

Квантовые точки и наночастицы на основе золота или серебра начали использовать в аналитической химии в конце XX в. В XXI в. были изобретены графены, а затем и другие двумерные наноструктуры. Первая статья по графенам была опубликована в журнале Science в октябре 2004 г.

Она имеет лучшее цитирование среди всех статей Nature и Science за рассматриваемый период. Судя по публикациям в Analytical Chemistry и других журналах Американского химического общества, графены используют в самых разных областях аналитической химии [7]. Синтез графенов впервые осуществили московские физики, работающие в Манчестерском университете — Новоселов и Гейм. В табл. 2 графены включены в раздел "Физика".

Из статей, никак не связанных с аналитическим приборостроением, лучшее цитирование имеют статьи по солнечной энергетике. Прорыв в этой области вызван использованием редких природных минералов перовскитов. Солнечная энергетика может в разы уменьшить себестоимость опреснения морской воды, что крайне актуально, например, для Саудовской Аравии. Статьи по экологии не являются статьями по экоаналитике — используется биометрия.

ЗАКЛЮЧЕНИЕ

Проанализирован перечень наиболее цитируемых статей в Nature и Science, по 10 за каждый год. Установлена значимая корреляционная связь между уровнем цитирования наиболее цитируемой и 10-й по цитированию статьей по каждому году. Этот результат теоретически обоснован свойствами порядковых статистик. Динамика цитирования статьи из данного перечня описывается закономерностями, аналогичными рассмотренным ранее для больших групп высоко-, средне- и малоцитируемых статей из журналов Американского химического общества. Зависимость представляет собой произведение логистической кривой Ферхюльста, что характерно для широкого круга популяционных процессов, и линейной убывающей функции, моделирующей падение интереса к научной информации (моральное старение). Численные коэффициенты, характеризующие временные особенности динамики и масштаб публикационной активности, иные. Характерный масштаб цитирования составляет порядка 1000 цитирований, характерное время морального старения — около 15 лет.

Распределение публикаций по областям знаний показало существенное различие по 3 временным диапазонам: 2001–2006, 2007–2012 и 2013–2018 гг. Основные изменения связаны с возрастающей долей публикаций по солнечным батареям за счет других разделов физики, устойчивое убывание доли публикаций по молекулярной биологии и колебательный характер публикационной

активности по химии, биологии и конкретно по ДНК и РНК.

Работа выполнена в рамках Государственного задания № 075-00780-19-00.

СПИСОК ЛИТЕРАТУРЫ

1. *Архипов Д.Б., Буляница А.Л., Щербаков А.П.* Вебометрический анализ и его использование для изучения тенденций развития аналитического приборостроения // Научное приборостроение. 2014. Т. 24, № 2. С. 52–60. URL: <http://iairas.ru/mag/2014/abst2.php#abst7>
2. *Van Noorden R., Maher B., Nuzzo R.* The top 100 papers // Nature. 2014. Vol. 514, no. 7524. P. 550–553. DOI 10.1038/514550a
3. *Sanger F., Nicklen S., Coulson A.R.* DNA sequencing with chain-terminating inhibitors // Proc. Natl. Acad. Sci. USA. 1977. Vol. 74, no. 12. P. 5463–5467.
4. *Margulies M., Egholm M., Altman W.E. et al.* Genome sequencing in microfabricated high-density picolitre reactors // Nature. 2005. Vol. 437. P. 376–380. DOI: 10.1038/nature03959
5. *Hindson B.J., Ness K.D., Masquelier D.A. et al.* High-throughput droplet digital PCR system for absolute quantification of DNA copy number // Anal. Chem. 2011. Vol. 83, no. 22. P. 8604–8610. DOI: 10.1021/ac202028g
6. *Watson J.D.* The human genome project: past, present, and future // Science. 1990. Vol. 248, no. 4951. P. 44–49. DOI: 10.1126/science.2181665
7. *Novoselov K.S., Geim A.K., Morozov S.V., Jiang D., Zhang Y., Dubonos S.V., Grigorieva I.V., Firsov A.A.* Electric field effect in atomically thin carbon films // Science. 2004. Vol. 306, no. 5696. P. 666–669. DOI: 10.1126/science.1102896

**Институт аналитического приборостроения РАН,
Санкт-Петербург**

Контакты: Буляница Антон Леонидович,
antbulyan@yandex.ru

Материал поступил в редакцию 20.05.2019

ANALYTICAL INSTRUMENTATION IN THE JOURNALS "NATURE" AND "SCIENCE" FOR 2001–2017. WEBOMETRIC ANALYSIS

D. B. Arkhipov, A. L. Bulyanitsa, A. P. Shcherbakov

Institute for Analytical Instrumentation of RAS, Saint Petersburg, Russia

A webometric analysis of the journals Nature and Science for 2001–2017 was carried out. For each year selected 10 most cited articles. The number of links was determined by the Web of Science and Google Scholar. It was found that, of 170 articles, at least 85 are directly related to analytical instrumentation. The formal selection criterion was the presence in the article of the mention of any analytical instrument (for example, a sequencer). The dynamics of citing articles was also analyzed and an easily interpreted mathematical model of the citation process was proposed. The time dependence of the number of citations is adequately described by the product of the logistic dependence and a linearly decreasing function. The first factor is characteristic of any population process with restrictions. The second factor can be interpreted as the effect of the obsolescence of scientific information. The characteristic time interval for achieving zero citation for the group of publications in question is significantly less than for medium- and low-quoting articles in journals of the American Chemical Society

Keywords: webometric analysis, highly cited publications, analytical instrumentation, citation dynamics, mathematical model, number of citations

Fig. Dynamics of annual number of citings Article [4].
On the abscissa axis — the year after publication, on the
ordinate axis — the number of references

Table 1. The distribution of the citation level by year for
two selected groups of publications (the most cited and the
10th by citation)

Table 2. Distribution of the number of articles by subject
for different time periods

REFERENCES

1. Arkhipov D.B., Bulyanitsa A.L., Scherbakov A.P. [Webometrical analysis and its use for study of analytical instrumentation development trends]. *Nauchnoe Priborostroenie* [Scientific Instrumentation], 2014, Vol. 24, no. 2, pp. 52–60.
URL: <http://iairas.ru/en/mag/2014/abst2.php#abst7>
2. Van Noorden R., Maher B., Nuzzo R. The top 100 papers // *Nature*. 2014. Vol. 514, no. 7524. P. 550–553. DOI: 10.1038/514550a
3. Sanger F., Nicklen S., Coulson A.R. DNA sequencing with chain-terminating inhibitors // *Proc. Natl. Acad. Sci. USA*. 1977. Vol. 74, no. 12. P. 5463–5467.
4. Margulies M., Egholm M., Altman W.E. et al. Genome sequencing in microfabricated high-density picolitre reactors // *Nature*. 2005. Vol. 437. P. 376–380. DOI: 10.1038/nature03959
5. Hindson B.J., Ness K.D., Masquelier D.A. et al. High-throughput droplet digital PCR system for absolute quantification of DNA copy number // *Anal. Chem*. 2011. Vol. 83, no. 22. P. 8604–8610. DOI: 10.1021/ac202028g
6. Watson J.D. The human genome project: past, present, and future // *Science*. 1990. Vol. 248, no. 4951. P. 44–49. DOI: 10.1126/science.2181665
7. Novoselov K.S., Geim A.K., Morozov S.V., Jiang D., Zhang Y., Dubonos S.V., Grigorieva I.V., Firsov A.A. Electric field effect in atomically thin carbon films // *Science*. 2004. Vol. 306, no. 5696. P. 666–669. DOI: 10.1126/science.1102896

Contacts: *Bulyanitsa Anton Leonidovich*,
antbulyan@yandex.ru

Article received by the editorial office on 20.05.2019