

ВРЕМЕННАЯ ДИНАМИКА ИНДЕКСА ХИРША

Ю.Ю. Тарасевич, Т.С. Шиняева

Проведен анализ информации из базы данных Scopus о временной зависимости индекса Хирша (h -индекса) и его модификации $h_s(2015)$ -индекса группы продолжительно и стабильно работающих ученых. Обнаружено, что характер изменения со временем $h_s(2015)$ -индекса близок к сигмоидальному. Предложена модель, описывающая динамику индекса Хирша. Модель учитывает: 1) изменение публикационной активности ученого – предполагается сигмоидальный рост числа публикаций на начальной стадии научной карьеры; 2) распределение статей по числу цитирований; 3) динамику цитирования каждой конкретной статьи (принято во внимание, что в типичном случае число цитирований сначала возрастает, а затем плавно убывает). Исследована динамика индекса Хирша в зависимости от средней продуктивности (числа публикуемых в течение года статей). Используются два вида распределения числа статей по числу цитирований: распределение Лотки и геометрическое распределение. Оба модельных распределения приводят к качественно верной временной динамике индекса Хирша.

Ключевые слова: моделирование; индекс Хирша.

Введение

Индекс Хирша или h -индекс – количественная характеристика публикационной активности ученого [1]. h -индекс ученого равен N , если N его публикаций имеют по N и более цитирований, а остальные публикации имеют не более N цитирований. h -индекс и его различные модификации, например, учитывающие нескольких соавторов у публикации [2, 3] или время научной активности ученого [1], широко используются в наукометрии (см. обзор и сравнение многочисленных модификаций h -индексов и родственных показателей [4]).

Индекс Хирша ученого меняется со временем. Хирш предположил, что h -индекс с годами растет приблизительно линейно [1]. Это предположение было подтверждено в рамках стохастической модели [5]. Однако, исследования наукометрических данных свидетельствуют, что линейный рост h -индекса – только одна из возможностей его временной динамики [6, 7]. Линейный рост может замедляться вплоть до выхода h -индекса на постоянное значение; возможно и более сложное поведение [6]. В рамках концепции процесса производства информации было показано, что $h = T^{1/\alpha}$, где T – полное число источников (публикаций), α – показатель Лотки [8]. В [9, 10] была предложена модель, в которой рост h -индекса с течением времени замедляется, и h -индекс выходит на постоянное значение. В [11] проведено компьютерное моделирование динамики изменения h -индекса со временем для различных моделей распределения цитирований. В большинстве случаев модель продемонстрировала линейный рост со временем h -индекса, в редких случаях наблюдалось замедление роста h -индекса.

Разнообразный характер зависимости h -индекса от времени [6, 7] заставляет задуматься о двух возможных альтернативах.

- Универсальное поведение h -индекса от времени отсутствует. Для каждого ученого динамика h -индекса уникальна.

- Динамика h -индекса является универсальной. В случае конкретного ученого эта динамика может не проявляться из-за низкой публикационной активности или ее незначительной продолжительности или быть замаскирована разнообразными факторами (например, радикальное изменение области научных исследований).

Целью настоящей работы является построение модели, которая позволила понять основные закономерности динамики h -индекса. В разделе «Экспериментальные данные» мы представляем результаты обработки данных о динамике h -индекса группы активно публикующихся ученых, представленных в базе данных Scopus. В разделе «Модель» описана математическая модель временной динамики h -индекса.

1. Экспериментальные данные

Для анализа была использована база данных Scopus. Были проанализированы наукометрические данные 25 ученых. Среди них – лауреаты Нобелевской премии по физике Ж. Алферов (2000), В. Гинзбург (2003), А. Гейм (2010), К. Новоселов (2010) и лауреат Филдсовской премии С. Смирнов (2010). Для анализа отбирались те ученые, про которых было достоверно известно, что они активно и стабильно публикуют результаты своих исследований в течение продолжительного времени. Кривые зависимости h -индекса от времени не демонстрируют общих закономерностей (рис. 1), что согласуется с проведенными ранее исследованиями [6, 7].

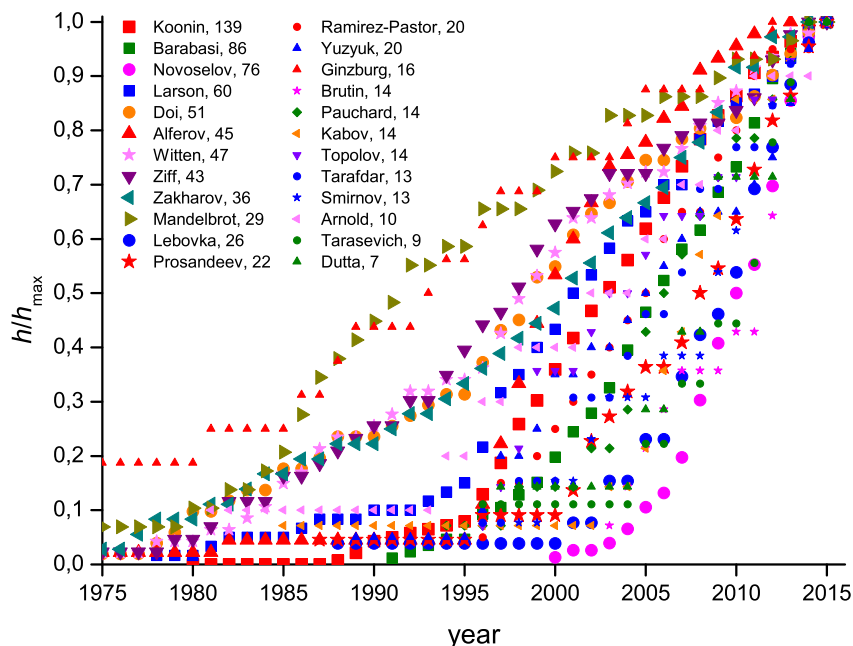


Рис. 1. Зависимость величины нормированного h -индекса от времени. После фамилии ученого через запятую указано значение его h -индекса по состоянию на 30 апреля 2015 года

Лишь некоторые кривые имеют почти линейные участки, соответствующие гипотезе Хирша [1], в большинстве случаев кривые вогнутые или выпуклые.

Кроме того, для каждого ученого была исследована динамика $h_s(t)$ -индекса. $h_s(t)$ -индекс – значение h -индекса в году t , который определяется работами, опубликованными до года s включительно [12]. Например, $h_{2000}(2015)$ – значение h -индекса данного автора в 2015 году, определенное на основании цитирований его работ, опубликованных до 2000 года включительно. Естественно, что $h_t(t)$ совпадает с обычным h -индексом. В отличие от обычного h -индекса, зависимость $h_s(t)$ -индекса для группы рассмотренных нами ученых близка к сигмоидальной. Зависимость величины h_s -индекса от времени аппроксимировалась функцией

$$h_s(t) = \frac{h_{\max} \exp\left(\frac{t-t_0}{\Delta t}\right)}{1 + \exp\left(\frac{t-t_0}{\Delta t}\right)}.$$

В качестве примера см. рис. 2.

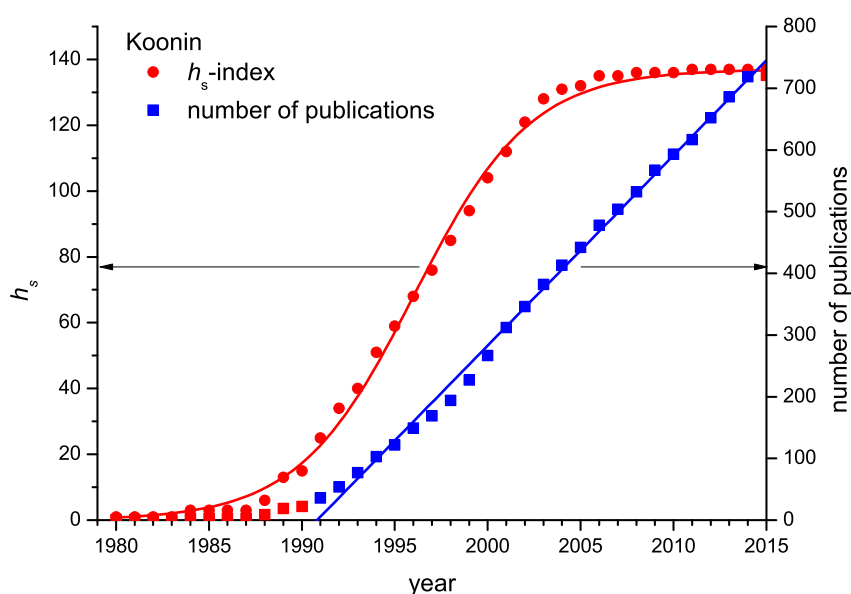


Рис. 2. Зависимость от времени h_s -индекса и общего числа публикаций и результаты аппроксимации исходных данных. В случае аппроксимации h_s -индекса $R^2 = 0.99797$, в случае аппроксимации числа публикаций $R^2 = 0.99622$. По данным проекта «Корпус экспертов по естественным наукам» (<http://expertcorps.ru/>) Е. Кунин имеет наибольший h -индекс среди российских ученых

В единственном случае (А. Гейм) была обнаружена зависимость h_s -индекса от времени, резко отличающаяся от остальных наборов данных (рис. 3). Мы связываем такое поведение h_s -индекса с началом работ по тематике графена. Для сравнения на том же рисунке представлены данные по К. Новоселову – соавтору А. Гейма, который к моменту начала «графенового бума» имел меньший период активной научной работы и меньшие значения наукометрических показателей.

Поскольку продолжительность активной работы ученых и максимальные значения h_s -индекса в выборке сильно отличались (h_s -индекс изменяется от 7 до 137), данные для последующего анализа были отмасштабированы: для каждого набора данных h_s -индекс поделили на h_{\max} , время – на Δt , кроме того, начало отсчета времени смещалось к значению t_0 . Данные представлены на рис. 4.

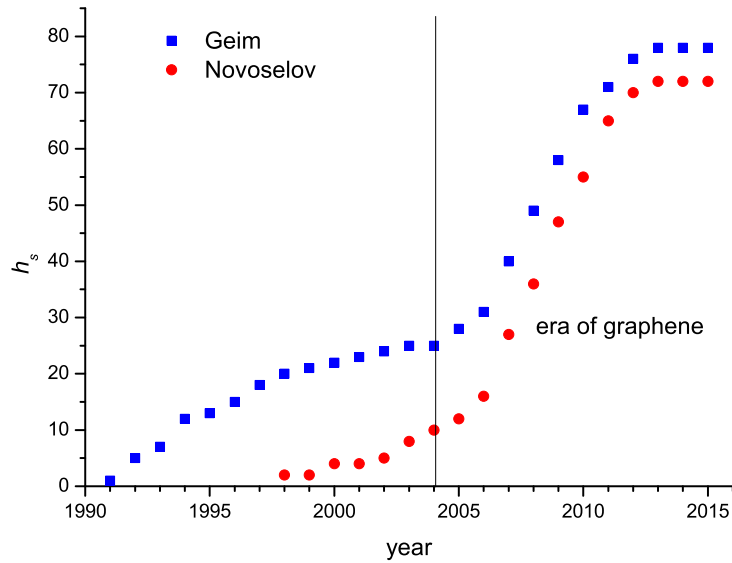


Рис. 3. Зависимость от времени h_s -индексов Нобелевских лауреатов А. Гейма и К. Новоселова

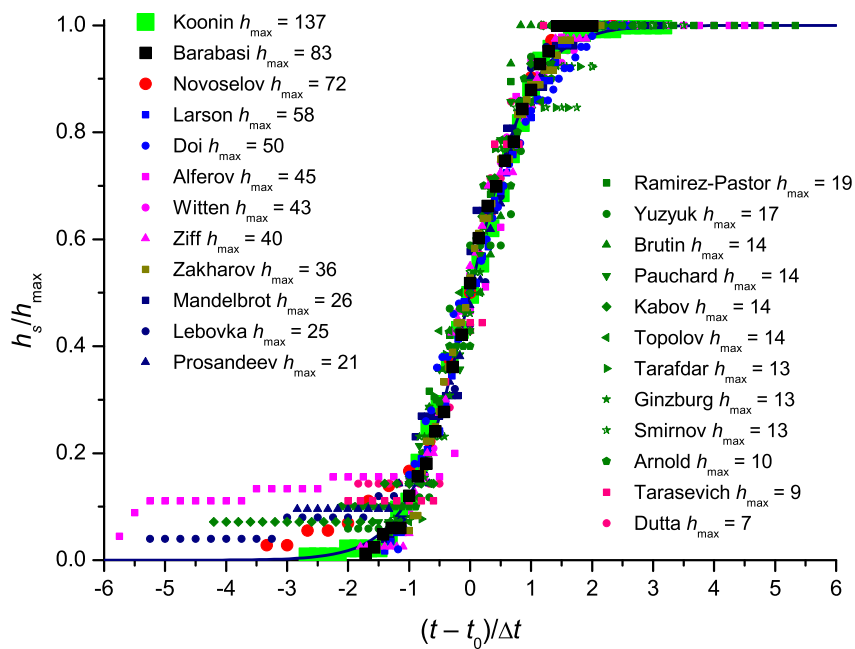


Рис. 4. Зависимость величины нормированного h_s -индекса от безразмерного времени по состоянию на 7 февраля 2015 года

Естественно, чем меньше значение h_s/h_{max} , тем заметнее дискретность h_s . Если для ученых с большими значениями h_{max} графики гладкие, то для ученых с малыми значениями h_{max} графики ступенчатые. Во всех рассмотренных случаях имеется временной участок с почти линейным ростом h_s -индекса. Снижением скорости роста h_s -индекса и его выход на постоянное значение отражает тот факт, что недавно опубликованные работы не оказывают влияния на величину h -индекса. Начальные участки зависимостей h_s -индекса демонстрируют в некоторых случаях заметное от-

клонение от сигмоидальной кривой. Эффект может быть объяснен, во-первых, существенным влиянием дискретности при малых значениях h_s , во-вторых, возможной неполнотой данных, поскольку относительно полная информация о цитированиях накоплена только после 1995 года (например, начало публикационной активности Ж. Алфéroва относится к 1973 году).

2. Модель

Временная динамика h -индекса может зависеть от нескольких факторов.

1. Временная динамика h -индекса связана с количеством статей, которые ученый публикует в течение года. Для активно работающего ученого эта величина не является постоянной: в начале научной карьеры ежегодное число публикаций растет, затем выходит на насыщение, в конце научной карьеры плавно (в случае постепенного отхода от дел) или резко (в случае смерти ученого) сокращается до нуля.
2. Временная динамика h -индекса связана с распределением статей ученого по числу цитирований. Количество цитирований косвенно связано с импакт-фактором журналов, в которых публикует свои работы ученый. Для активно работающего ученого разумно ожидать, что средняя величина импакт-фактора журналов, в которых публикуется ученый, растет в течение научной карьеры ученого. Можно ожидать, что распределение статей ученого по цитированиям при $t \rightarrow \infty$ зависит от среднего импакт-фактора журналов, в которых опубликованы статьи. Так, для статей по биологии, химии и биохимии была найдена линейная связь между числом ссылок на статью и импакт-фактором журнала, в котором она была опубликована; для статей по математике и физике эта связь почти линейная [13]. Таким образом, ученый, который публикует свои результаты в журналах с высоким импакт-фактором, будет иметь большую долю высоко цитируемых статей.
3. Временная динамика h -индекса связана с динамикой цитирования каждой конкретной статьи.
4. Количество цитирований статьи линейно зависит от числа соавторов и числа ссылок [13]. Поскольку ученые обычно отслеживают публикации известных им коллег, увеличение числа соавторов приводит к увеличению вероятности того, что публикация будет замечена и использована. Кроме того, обычно ученые отслеживают ссылки на некоторые важные для них публикации. Увеличение списка цитированных источников, таким образом, увеличивает вероятность того, что работу заметят. Поскольку h -индекс зависит от числа цитирований, перечисленные факторы оказывают опосредованное влияние на динамику h -индекса.

В нашей модели мы постарались учесть первые три из перечисленных эффектов. Исследованию последнего эффекта посвящены, например, работы [14, 15].

Для проведения моделирования необходимо знать, как меняется число цитирований конкретной статьи от времени. Число цитирований статьи в единицу времени меняется с годами и зависит от типа статьи [16]. В [16] предложена кривая, описывающая динамику числа ссылок на статью. Эта кривая имеет максимум, положение

и высота которого зависит от типа статьи (обычная исследовательская статья, письмо редактору, обзор). Типичные кривые для различных типов статей приведены на рис. 5. По форме кривая, описывающая динамику цитирований, близка к логнормальному распределению [17]

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right). \quad (1)$$

Заметим, что $x_c = e^\mu$ является медианой и соответствует широко используемой в наукометрии величине cited half-life. Для статей по физике эта величина приблизительно равна 7 годам, по химии и техническим наукам – около 8 лет [18]. Поскольку $\int_0^\infty f(x) dx = 1$, то $f(x)$ следует умножать на полное число цитирований i -й публикации за бесконечно большое время c_i^{\max} . В [11] использован для моделирования упрощенный закон изменения цитирований с плотностью распределения в виде разностороннего треугольника. Для описания динамики затухания цитирований статьи (стадия после максимума цитирования) предложена иная математическая формула [19], однако мы будем опираться на формулу (1), поскольку нас интересуют все стадии цитирования статьи.

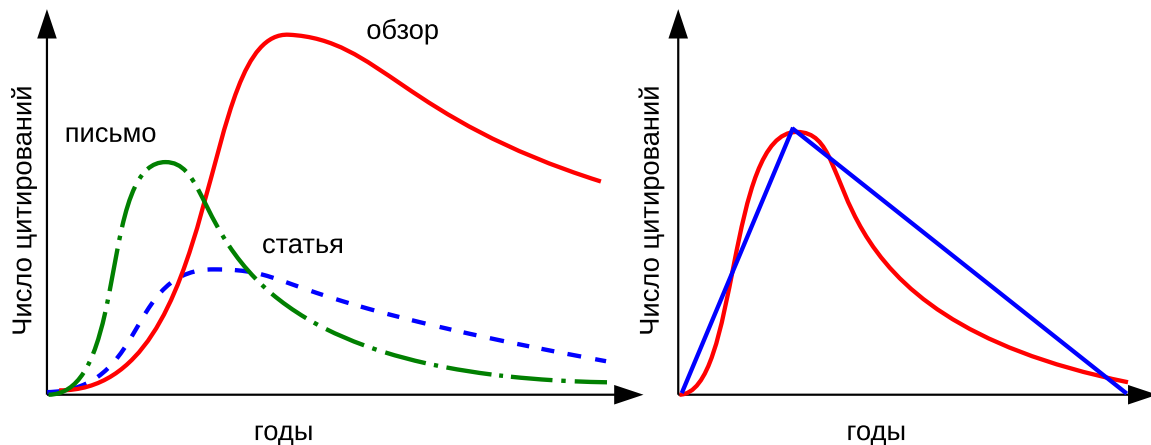


Рис. 5. Слева: типичные зависимости от времени числа цитирований различных типов научных публикаций [16]. Справа: типичная зависимость числа цитирований от времени (кривая) и ее приближенное представление (ломанная) [11]

Понятно, что конкретная статья цитируется лишь приближенно в соответствии с теоретической кривой. При моделировании для генерации числа цитирований статьи в момент времени t используем распределение Пуассона

$$p_k = \frac{\lambda^k}{k!} e^{-\lambda} \quad (2)$$

с параметром $\lambda = c_i^{\max} f(t)$ (рис. 6).

Для конкретной статьи cited half-life может сколь угодно сильно отличаться от среднего значения. Например, в течение почти 40 лет число ежегодных цитирований статьи (Scopus eid=2-s2.0-0001155449) продолжает расти, таким образом cited half-life

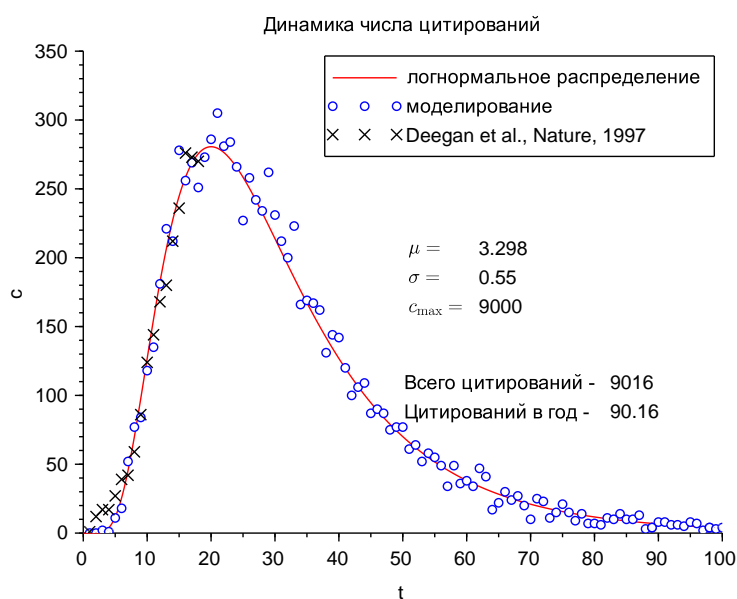


Рис. 6. Число цитирований некоторой публикации c как функция времени, прошедшего с момента публикации: теоретическая кривая (1), моделирование с использованием распределения (2) и сравнение с реальной динамикой цитирования статьи по данным Scopus (eid=2-s2.0-0030732736)

для этой статьи заведомо превышает 40 лет, отличаясь от типичного значения в 6 или более раз (см. рис. 7); данные о цитированиях статьи (Scopus eid=2-s2.0-0030732736, рис. 6) также демонстрируют, что cited half-life существенно превосходит типичное значение.

Распределению публикаций по числу цитирований посвящены многочисленные работы. Обзор подходов, применяющихся при моделировании распределения публикаций по числу цитирований, с соответствующими ссылками можно найти в [20]. Экспериментальное установление реального закона распределения или проверка некоей теоретической гипотезы невозможно, поскольку такая проверка предполагает наличие полных данных за весьма длительный (теоретически – бесконечный) промежуток времени после публикации, когда цитирования статей полностью прекращаются. Однако, относительно полные данные накоплены только о цитированиях статей после 1995 года, т.е. за 20 последних лет, а научные работы нередко цитируются многие десятилетия (рис. 7).

Зачастую считается, что при $t \rightarrow \infty$ распределение публикаций по числу цитирований соответствует модели Лотки [21]

$$\varphi(x) = \frac{\alpha - 1}{x^\alpha}, \tag{3}$$

где $\varphi(x)$ – функция вероятности.

На основании данных Web of Science о цитированиях более 200000 статей по математике, физике, химии, биохимии и биологии было показано [13], что реальное распределение публикаций по числу цитирований хорошо аппроксимируется геомет-

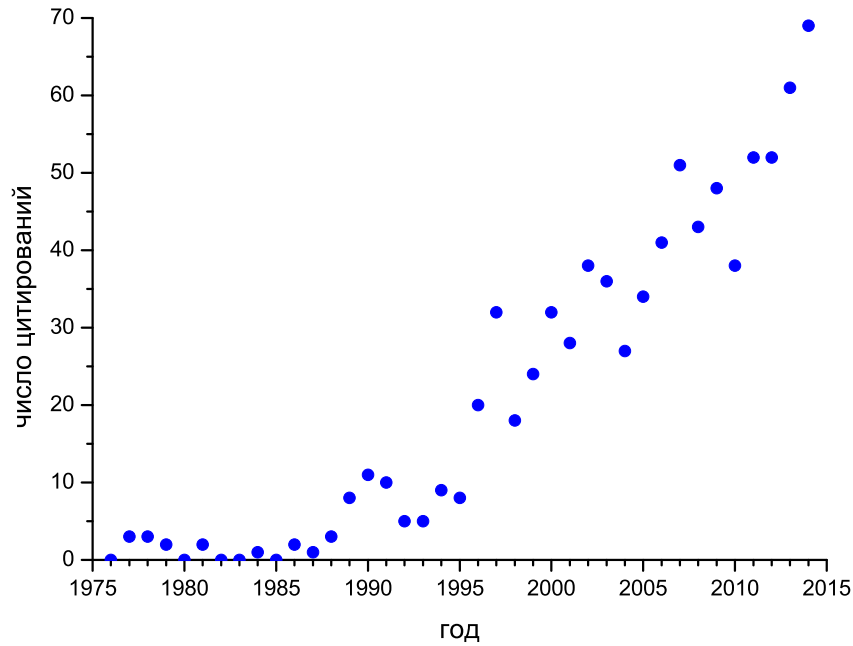


Рис. 7. Число цитирований статьи по данным Scopus (eid=s2.0-0001155449)

рическим распределением

$$p_k = \frac{1}{E + 1} \left(\frac{E}{E + 1} \right)^k, \quad (4)$$

где p_k – вероятность того, что статья будет процитирована k раз, k – ожидаемое число цитирований статьи, E – параметр. Для статей по физике, например, $E = 9,72$. Точность аппроксимации существенно улучшается, если использовать сумму двух геометрических распределений с разными параметрами. Следует иметь в виду, что проведенный анализ основан на данных о цитированиях за ограниченный промежуток времени, в пределе $t \rightarrow \infty$ результаты могут быть иными. Кроме того, при рассмотрении распределения статей конкретного ученого по цитированиям можно ожидать, что величина E зависит от уровня ученого – у выдающихся ученых E велико, у слабых – мало. Соответственно, у сильного ученого доля статей с малым числом цитирований меньше, зато имеются статьи с очень большим числом цитирований; у слабого ученого, напротив, доминируют статьи с нулевым или малым числом цитирований, высоко цитируемые статьи не встречаются. Тот факт, что данные по большому массиву статей хорошо аппроксимируются комбинацией двух геометрических распределений, может свидетельствовать о том, что в обрабатываемой совокупности представлены статьи, написанные авторами разного уровня.

Алгоритм

1. Задаем время активной деятельности ученого. В приведенных в дальнейшем результатах этот параметр всюду равен 50. Задаем среднее число статей $\langle n \rangle$, которые ученый публикует за год. Считаем, что публикационная активность с годами меняется – предполагается сигмоидальный рост числа публикаций на начальной стадии научной карьеры. Принимая в качестве нулевого приближения,

что на протяжении всей своей научной деятельности ученый работает с одинаковой активностью, то есть в среднем публикует ежегодно одно и то же число статей, оцениваем полное число статей, опубликованных за все время научной деятельности, как $N = \langle n \rangle t_{\max}$.

2. Число статей $n(t)$, опубликованных в год t , генерируем случайным образом по закону Пуассона с $\lambda = \langle n \rangle$. Таким образом, в среднем за год будут появляться $\langle n \rangle$ статей, но в конкретном году число статей может быть больше или меньше.
3. В зависимости от выбранной модели распределения публикаций по числу цитирований с помощью распределения Лотки (3) или геометрического распределения (4) создаем последовательность N случайных чисел c_i^{\max} – полное число цитирований i -й статьи за все время. Первые $n(1)$ чисел из этой последовательности относятся к статьям, опубликованным в 1 год, следующие $n(2)$ чисел относятся к статьям, опубликованным во второй год, и т.д.
4. С помощью формулы (1) получаем распределение ожидаемого числа цитирований i -й статьи по годам. Параметры логнормального распределения для каждой статьи меняются в зависимости от максимального ожидаемого числа ссылок на статью. Логично предположить, что чем больше ссылок на статью, тем дольше продолжаются цитирования, и тем больше время, когда цитирования достигают максимума. Мы положили, что моде M соответствует время $\sqrt{c_i^{\max}}$, а cited half-life в полтора раза больше $T_{1/2} = 1,5M$. Тогда, $\mu = \ln T_{1/2}$, $\sigma^2 = \ln \left(\frac{T_{1/2}}{M} \right)$, что обеспечивает близость ожидаемой динамики цитирований к наблюдаемой. Таким образом, для i -й статьи ожидается число цитирований в год $t > t_i$, где t_i – год, в который была опубликована статья,

$$f_i(t) = \frac{c_i^{\max}}{(t - t_i)\sigma_i\sqrt{2\pi}} \exp \left(-\frac{(\ln(t - t_i) - \mu_i)^2}{2\sigma_i^2} \right).$$

5. Для года t генерируем случайное число $c_i(t)$ – число цитирований i -й статьи в году t . Случайные числа должны соответствовать распределению Пуассона. Математическое ожидание числа цитирований i -й статьи в год t равно $\lambda = f_i(t)$.

3. Результаты моделирования

Моделирование проводилось для двух распределений публикаций по числу цитирований. Использовалось распределение Лотки (3) с различными показателями α , геометрическое распределение (4), а также сумма двух геометрических распределений, параметры которых были определены по методу наименьших квадратов для данных по физическим журналам, представленным в [13]. Динамика h -индекса даже при одинаковых параметрах демонстрирует заметный разброс, поэтому представленные в настоящей работе результаты были получены путем усреднения серии из 10 экспериментов при каждом наборе параметров.

Результаты моделирования с использованием распределения Лотки для распределения статей по числу цитирований представлены на рис. 8. Уменьшение показателя в распределении Лотки приводит к увеличению максимального значения h -индекса при любой продуктивности.

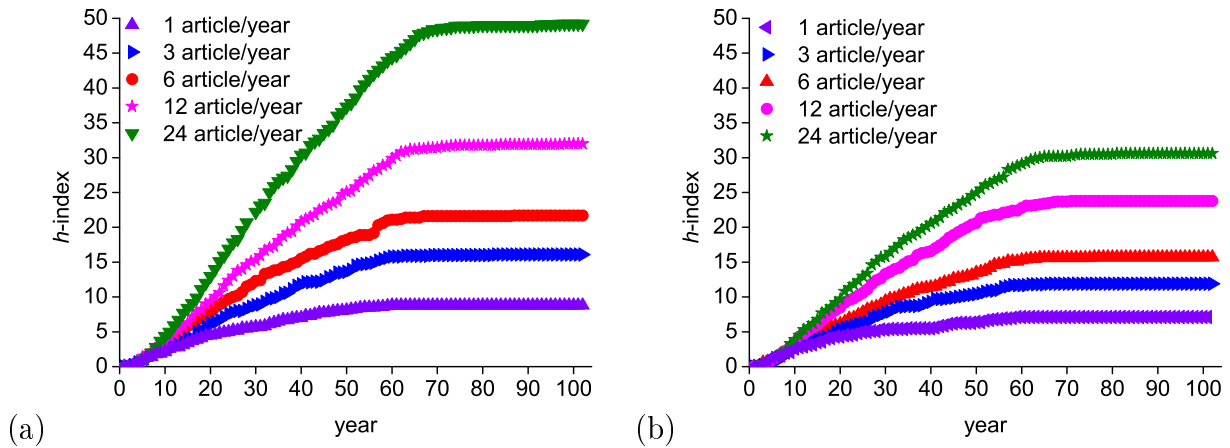


Рис. 8. Результаты моделирования динамики h -индекса при различном числе публикаций в год с использованием распределения статей по числу цитирований в соответствии с распределением Лотки, (a) $\alpha = 1.8$, (b) $\alpha = 2.0$. Каждая кривая получена путем усреднения результатов по 10 реализациям

Результаты моделирования с использованием геометрического распределения представлены на рис. 9.

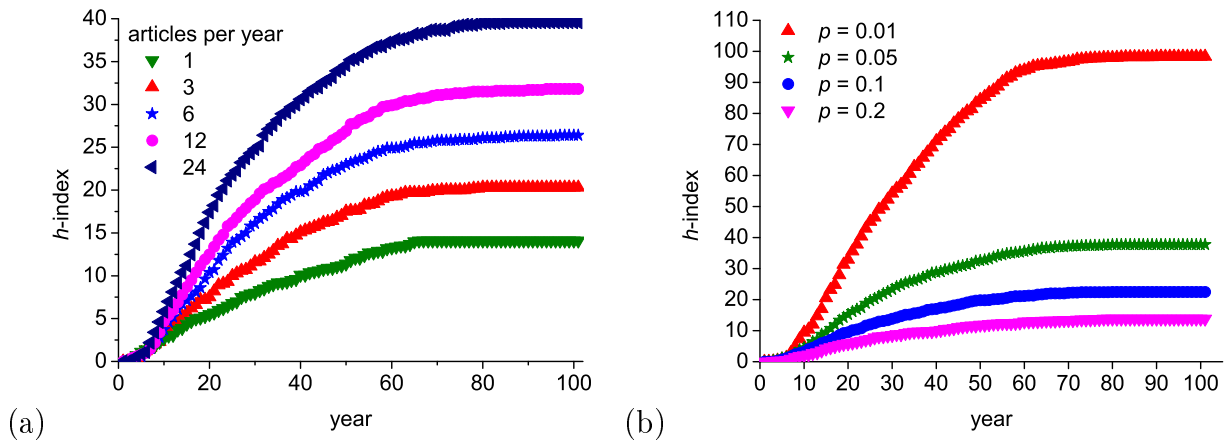


Рис. 9. Результаты моделирования динамики h -индекса: а) при различной продуктивности ученого (среднем числе публикаций в год) с использованием распределения статей по числу цитирований в соответствии с данными [13]; б) при средней продуктивности ученого 6 публикаций в год с использованием распределения статей по числу цитирований в соответствии с геометрическим распределением при разных значениях параметра распределения, $p = (1 + E)^{-1}$. Каждая кривая получена путем усреднения результатов по 10 реализациям

Для обеих гипотез о характере распределения публикаций по числу цитирований кривые зависимости h -индекса от времени качественно согласуются с данными о динамике изменения h -индекса реальных ученых. Имеется согласие результатов моделирования с реальными данными и в случае h_s -индекса (рис. 10).

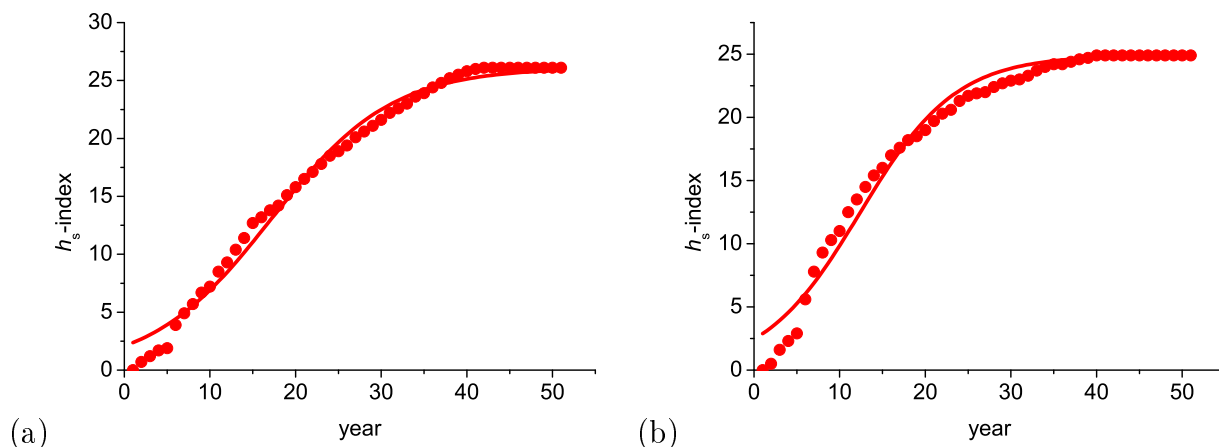


Рис. 10. Результаты моделирования изменения h_s -индекса при продуктивности 12 статей в год. Распределение статей по числу цитирований соответствует (a) распределению Лотки, $\alpha = 1, 8$; (b) геометрическому распределению, $p = 0, 1$

В нынешнем виде модель использует два упрощающих предположения. Отказ от этих предположений может обеспечить лучшее согласие результатов моделирования с реальными данными.

1. Предположение о том, что распределение статей по числу цитирований одинаково для всех ученых. Разумно предположить, что высокая продуктивность коррелирует с качеством публикаций. Вероятно, большое число публикаций типично для руководителей крупных лабораторий. Стать руководителем научной группы может ученый, имеющий вес в научном сообществе. Логично предположить, что у более продуктивного ученого по сравнению со средним ученым больше доля статей с большим числом цитирований и меньше доля статей с малым и нулевым числом цитирований.
2. Предположение о том, что качество публикаций ученого с годами не меняется. Можно ожидать, что с годами средний уровень публикаций растет.

Заключение

Нужно очень отчетливо осознавать, что наукометрические показатели в принципе не могут оценивать научный уровень получаемых ученым результатов. Наукометрические показатели характеризуют степень интеграции ученого в процесс обмена информацией внутри международного научного сообщества. Удостоенный Филдсовской премии Г. Перельман в базе данных Scopus вообще не представлен.

Следует учитывать, что число цитирований и динамика изменения числа цитирований зависят от числа соавторов публикации, а также от их известности в научных кругах [22]. Интуитивно понятно, что шансов привлечь к себе внимание при прочих равных условиях больше у той статьи, среди соавторов которой имеется известный ученый.

Можно было бы ожидать, что, имея данные о динамике h -индекса на среднем этапе научной карьеры ученого, можно прогнозировать динамику дальнейшего роста h -индекса. Такой прогноз можно было бы рассматривать как надежный, поскольку

h -индекс является весьма инерционным показателем и зависит от всей предыдущей публикационной активности ученого [23].

h -индекс – интегральная величина, что одновременно является его достоинством и недостатком. Простота вычисления и интерпретации h -индекса делают его удобной количественной характеристикой. Интегральный характер h -индекса делает его непригодным для детального исследования публикационной активности ученого.

Индекс Хирша не является в чистом виде индивидуальным показателем ученого, т.к. большинство работ в настоящее время публикуется в соавторстве, а количество цитирований определяется не только научным уровнем публикации, но и известностью в научных кругах авторов, количеством ученых, работающих по тематике публикации, а также рядом других факторов, которые вообще трудно учесть.

В наши дни наука все более становится публичной деятельностью, и для нее оказываются справедливыми те же самые закономерности, что и для шоу-бизнеса. Если оценивать по количеству упоминаний в СМИ оперных певцов и певцов других жанров, то первые окажутся в проигрыше, что отнюдь не свидетельствует о незначительности вклада оперного искусства в мировую культуру.

При использовании любого формального критерия для оценки научной деятельности следует проявлять максимальную осторожность. Даже внутри весьма компактной и авторитетной группы ученых – действительных членов РАН – имеется весьма большой (десятикратный) разброс наукометрических показателей (рис. 11).

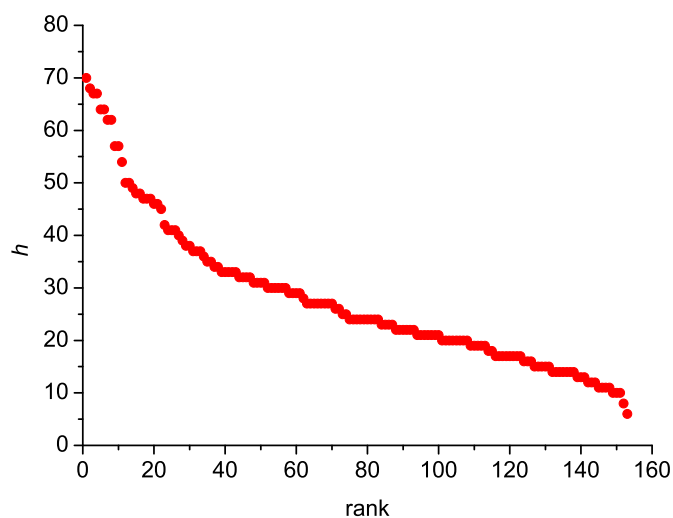


Рис. 11. Значения h -индекса академиков РАН, имеющих не менее 100 цитирований работ, опубликованных за последние 7 лет. По данным проекта «Корпус экспертов по естественным наукам» (<http://expertcorps.ru/>)

Юрий Юрьевич Тарасевич, доктор физико-математических наук, профессор, заведующий лабораторией «Математическое моделирование и информационные технологии в науке и образовании», Астраханский государственный университет (г. Астрахань, Российская Федерация), tarasevich@asu.edu.ru.

Таисия Сергеевна Шиняева, аспирант, Астраханский государственный университет, (г. Астрахань, Российская Федерация), danilova.taisya@gmail.com.

Поступила в редакцию 1 июня 2015 г.

Литература / References

1. Hirsch J.E. An Index to Quantify an Individual's Scientific Research Output. *Proceedings of the National Academy of Sciences of the United States of America*, 2005, vol. 102, no. 46, pp. 16569–16572. DOI: 10.1073/pnas.0507655102
2. Schreiber M. A Modification of The h -index: The h_m -index Accounts for Multi-Authored Manuscripts. *Journal of Informetrics*, 2008, vol. 2, no. 3, pp. 211–216. DOI: 10.1016/j.joi.2008.05.001
3. Hirsch J.E. An Index to Quantify an Individual's Scientific Research Output That Takes Into Account the Effect of Multiple Coauthorship. *Scientometrics*, 2010, vol. 85, no. 3, pp. 741–754. DOI: 10.1007/s11192-010-0193-9
4. Schreiber M. Twenty Hirsch Index Variants and Other Indicators Giving More or Less Preference to Highly Cited Papers. *Annalen der Physik*, 2010, vol. 522, no. 8, pp. 536–554. DOI: 10.1002/andp.201000046
5. Burrell Q.L. Hirsch's h -index: A Stochastic Model. *Journal of Informetrics*, 2007, vol. 1, no. 1, pp. 16–25. DOI: 10.1016/j.joi.2006.07.001
6. Burrell Q.L. Hirsch Index or Hirsch Rate? Some Thoughts Arising From Liang's Data. *Scientometrics*, 2007, vol. 73, no. 1, pp. 19–28. DOI: 10.1007/s11192-006-1774-5
7. Wu J. Empirical Study of the Growth Dynamics in Real Career h -index Sequences. *Journal of Informetrics*, 2011, vol. 5, no. 4, pp. 489–497. DOI: 10.1016/j.joi.2011.02.003
8. Egghe L., Rousseau R. An Informetric Model for the Hirsch-Index. *Scientometrics*, 2006, vol. 69, no. 1, pp. 121–129. DOI: 10.1007/s11192-006-0143-8
9. Egghe L. Dynamic h -index: The Hirsch Index in Function of Time. *Journal of the American Society for Information Science and Technology*, 2007, vol. 58, no. 3, pp. 452–454. DOI: 10.1002/asi.20473
10. Egghe L. Item-time-dependent Lotkaian Informetrics and Applications to the Calculation of the Time-Dependent h -index and g -index. *Mathematical and Computer Modelling*, 2007, vol. 45, no. 7-8, pp. 864–872. DOI: 10.1016/j.mcm.2006.08.006
11. Guns R., Rousseau R. Simulating Growth of the h -index. *Journal of the American Society for Information Science and Technology*, 2009, vol. 60, no. 2, pp. 410–417. DOI: 10.1002/asi.20973
12. Schreiber M. How Relevant is the Predictive Power of the h -index? A Case Study of the Time-Dependent Hirsch Index. *Journal of Informetrics*, 2013, vol. 7, no. 2, pp. 325–329. DOI: 10.1016/j.joi.2013.01.001
13. Vieira E.S, Gomes J. A. N. F. Citations to Scientific Articles: Its Distribution and Dependence on the Article Features. *Journal of Informetrics*, 2010, vol. 4, no. 1, pp. 1–13. DOI: 10.1016/j.joi.2009.06.002
14. Sangwal K. Growth Dynamics of Citations of Cumulative Papers of Individual Authors According to Progressive Nucleation Mechanism: Concept of Citation Acceleration. *Information Processing & Management*, 2013, vol. 49, no. 4, pp. 757–772. DOI: 10.1016/j.ipm.2013.01.003
15. Amancio D.R., Oliveira Jr. O.N., L. da Fontoura Costa. Three-Feature Model to Reproduce the Topology of Citation Networks and the Effects from Authors' Visibility on Their h -index. *Journal of Informetrics*, 2012, vol. 6, no. 3, pp. 427–434. DOI: 10.1016/j.joi.2012.02.005
16. Amin M., Mabe M.A. Impact Factors: Use and Abuse. *Medicina*, 2003, vol. 63, no. 4, pp. 347–354.

17. Egghe L., Ravichandra Rao I.K. Theory of First-Citation Distributions and Applications. *Mathematical and Computer Modelling*, 2001, vol. 34, no. 1-2, pp. 81–90. DOI: 10.1016/S0895-7177(01)00050-4
18. Tsay M.-Y. An Analysis and Comparison of Scientometric Data Between Journals of Physics, Chemistry and Engineering. *Scientometrics*, 2009, vol. 78, no. 2, pp. 279–293. DOI: 10.1007/s11192-007-1996-1
19. Bouabid H. Revisiting Citation Aging: a Model for Citation Distribution and Life-Cycle Prediction. *Scientometrics*, 2011, vol. 88, no. 1, pp. 199–211. DOI: 10.1007/s11192-011-0370-5
20. Sangwal K. Distributions of Citations of Papers of Individual Authors Publishing in Different Scientific Disciplines: Application of Langmuir-Type Function. *Journal of Informetrics*, 2014, vol. 8, no. 4, pp. 972–984. DOI: 10.1016/j.joi.2014.09.009
21. Egghe L. Mathematical Study of h -index Sequences. *Information Processing & Management*, 2009, vol. 45, no. 2, pp. 288–297. DOI: 10.1016/j.ipm.2008.12.002
22. McCarty C., Jawitz J.W., Hopkins A., Goldman A. Predicting Author h -index Using Characteristics of the Co-author Network. *Scientometrics*, 2013, vol. 96, no. 2, pp. 467–483. DOI: 10.1007/s11192-012-0933-0
23. Schreiber M. The Predictability of the Hirsch Index Evolution. Translational Twists and Turns: Science as a Socio-Economic Endeavor. *Proceedings of 18th International Conference on Science and Technology Indicators. Berlin, Germany, September 4–6, 2013*. Berlin, Institute for Research Information and Quality Assurance, 2013, pp. 366–372.

MSC 62P25

DOI: 10.14529/mmp160103

TEMPORAL DYNAMICS OF HIRSCH INDEX

Yu. Yu. Tarasevich, Astrakhan State University, Astrakhan, Russian Federation,
tarasevich@asu.edu.ru,

T. S. Shinyayeva, Astrakhan State University, Astrakhan, Russian Federation,
danilova.taisiya@gmail.com

We performed the analysis of the data from the Scopus database regarding temporal dynamics of h -index and $h_s(2015)$ -index of a group of the continuously and consistently working scientists. We propose a model describing the temporal dynamics of h -index. Temporal dynamics of $h_s(2015)$ -index demonstrates sigmoidal behaviour. The model takes into account: 1) changing the publication activity of the scientist (sigmoidal growth of number of publications at the early stages of scientific career is assumed); 2) the distribution of articles by the number of citations; 3) the dynamics of each specific article citation (typically, the number of citations at first increases and then gradually decreases). The dynamics of the h -index as a function of average productivity (number of articles published per year) is investigated. We used two types of citations distributions, i.e. Lotka distribution and geometric distribution. Both distributions lead to a qualitatively correct temporal dynamics of Hirsch index.

Keywords: h-index; modelling.

Received June 1, 2015