ФОРМАЛИЗАЦИЯ ИНДЕКСА ХИРША
Обзор

С. В. Бредихин, Н. Г. Щербакова

Институт вычислительной математики и математической геофизики СО РАН,
630090, Новосибирск, Россия

УДК 001.12+303.2

Представлены аксиоматическое определение индекса Хирша и его основных свойств; три теоретических модели, интерпретирующие этот индекс через параметры распределения цитирований, а также материал об h-последовательностях, позволяющих изучить динамику изменения h-индекса и расширить информацию о манере цитирования.

Ключевые слова: индекс Хирша, h-индекс для журналов, аксиоматика h-индекса, теоретические модели, h-последовательности.

The axiomatic characterization of the Hirsch-index and its main properties; three theoretical models of the dependence of the index on the distribution parameters of citations and the study of h-sequences that allow to learn the dynamics of index changings and to complement the absent details of citing manner are submitted.

Keywords: Hirsch index, journal h-index, the theoretic models, h-sequences.

Введение. Индекс Хирша, или h-индекс (англ. *h-index*), предложен Х. Хиршем в 2005 г. [1, 2] в качестве наукометрического показателя продуктивности ученого, основанного на распределении цитирования его работ. Определение h-индекса ученого: ученый имеет h-индекс, равный h, если h из его N_p публикаций имеют по крайней мере h цитирований каждая, а остальные $N_p - h$ публикаций имеют не более чем h цитирований каждая. Множество всех публикаций автора, удовлетворяющих этому определению, называют h-ядром. В случае соавторства при подсчете h-индекса Хирш предлагал делить количество цитирований работы на количество соавторов и сопоставлять полученную долю с каждым автором.

Введение h-индекса Хирш аргументировал тем, что этот критерий оценки предпочтительнее других подобных индексов, таких как количество работ, деленное на общее количество цитирований, или количество цитирований, приходящихся на одну работу. Однако Хирш считал, что применение одной этой количественной меры может дать только грубую аппроксимацию индивидуального профиля ученого, ее следует применять в случае, если дело касается грантов или подтверждения статуса ученого. В работе [3] отмечено, что преимущество h-индекса перед другими показателями подобного рода заключается в учете как количества публикаций, так и их востребованности (число цитирований этих публикаций). Таким образом, h-индекс является результатом баланса между количеством публикаций и количеством цитирований, полученных каждой публикацией. Например, он учитывает, что ученый (индивидуально или в соавторстве) может опубликовать одну или несколько выдающихся работ, которые будут много цитироваться, однако не обеспечат равномерную производительность автора. В то же время h-индекс „поощряет“ ученых, имеющих непрерывный поток публикаций и постоянную цитируемость или цитируемость выше средней нормы [4]. В. Гланзель [3] считает, что сила этого индекса — в потенциальной возможности оценивать небольшие наборы публикаций, к которым зачастую неприменимы традиционные библиометрические индикаторы.

Сравнение h-индекса с другими библиометрическими оценками проводилось многими авторами. Так в работе [5] подтверждается соответствие величины индекса оценке коллег. В работе [6] выявлена корреляция h-индекса с оценкой, полученной на основе подсчета общего количества цитирований. В работе [7] подтверждается корреляция h-индекса с несколькими библиометрическими показателями, а также с оценкой коллег. Обнаружена корреляция h-индекса с общим количеством публикаций [8]. Эти исследования подтверждают правомерность использования h-индекса.

Однако h-индекс имеет ряд недостатков, так авторы работы [9] отмечают, что недостатки в основном связаны с его неспособностью дифференцировать активных и неактивных ученых, выявлять важные работы, созданные в прошлом, а также работы, которые задают тенденцию и продолжают влиять на научное мышление. Сам Х. Хирш к недостаткам h-индекса относил его зависимость от продолжительности научной карьеры ученого и области исследований, имеющей свои традиции цитирования. В работе [10] высказывается сомнение в корректности применения h-индекса для вычисления научной эффективности, прежде всего обусловленное тем, что не учитывается вся история цитирования.

На основе h-индекса определено большое количество новых индексов, предназначенных для преодоления недостатков и использования совместно с h-индексом. Тем более что в работах ряда авторов высказывается предположение о невозможности использования одномерной метрики в многомерном пространстве библиометрии [3]. Предоставление h-индекса и его модификаций наукометрическими базами данных, такими, например, как *Web of Science* (*WoS*, *ThomsonReuters*) и *Scopus* (*Elsevier*), в качестве индикатора эффективности (менее чем через два года после определения!) является показателем того, что он стал общепринятой мерой академических достижений.

1. Индекс Хирша для журналов. Классическое определение индекса Хирша предназначено для сравнения эффективности научной работы авторов. В работе [11] приводится мнение, что широкому распространению h-индекса в качестве меры эффективности работы ученых мешают как недостатки (отсутствие стандартов цитирования в научных дисциплинах, неадекватное взвешивание соавторства и др.), так и естественное и обоснованное нежелание научного сообщества использовать для оценки численные индексы. Однако существуют области библиометрии, в которых меры, основанные на цитировании, получили большее признание. Одной из таких областей является анализ цитирования журналов.

В работах [11], [12] определяется аналог h-индекса для журналов (далее hJ). Индекс hJ вычисляется так. Пусть журнал J имеет N_p публикаций за рассматриваемый период времени T. Публикации упорядочим в порядке убывания цитирований. Из этой последовательности выберем h статей, имеющих по крайней мере h цитирований, так, чтобы каждая из оставшихся $N_p - h$ статей имела число цитирований, меньшее или равное h. В этом случае говорят, что журнал J за период T имеет индекс Хирша журнала hJ. Индекс выгодно дополняет индикаторы „важности“ журналов как мера, устойчивая к выбросам.

Как правило, индекс hJ применяется для оценки более широкого влияния, чем академическое, и рассматривается как мера социальной и экономической важности. Его следует с осторожностью использовать при учете индивидуальной производительности ученых, поскольку высокоцитируемые работы не обязательно печатаются влиятельными журналами, и наоборот, существенная часть работ, опубликованных в самых влиятельных журналах, не является высокоцитируемой.

2. Определения и аксиомы. Строгое определение h-индекса для автора приведено в работе [13]. Оно формулируется следующим образом. Предположим, что некоторый автор опубликовал n трудов, и каждая i-я публикация имеет X_i цитирований $(i = 1, 2, \ldots, n)$. Упорядочим публикации по „рангу“, т. е. по убыванию значений X_i

$$X_1^* \geq X_2^* \geq \ldots \geq X_n^*,$$

где X_1^* — число цитирований, полученное наиболее цитируемой публикацией; X_n^* — число цитирований, полученное наименее цитируемой публикацией. Будем считать, что значение h-индекса для автора вычисляется по формуле

$$h = \max\{j : X_j^* \ \geq \ j\},$$

в отличие от оригинального определения Х. Хирша, которое в этих терминах можно сформулировать следующим образом: такое число h, что

$$h \leq X_h^*, \text{ при этом } h > X_{h+1}^*,$$

в строгом определении решена проблема, когда $X_{h+1}^* = X_h^* = h$.

В работе [14] приводятся результаты исследования необходимых и достаточных условий, уникально характеризующих h-индекс. Условия сформулированы в виде аксиом, учитывающих изменения в количестве и качестве: количестве публикаций и цитируемости.

Исследователь, имеющий $n \geq 0$ публикаций, формально описывается с помощью вектора $\mathbf{x}= (x_1, x_2, \ldots, x_n)$ с неотрицательными компонентами $x_1 \geq x_2 \geq \ldots \geq x_n$; k-я компонента вектора определяет количество цитирований, полученных k-й работой; работы упорядочены по количеству цитирований. Если исследователь не имеет работ, вектор пуст. Пусть X — множество таких векторов. Будем говорить, что вектор $\mathbf{y}= (y_1, y_2, \ldots, y_m)$ доминирует над вектором $\mathbf{x}= (x_1, x_2, \ldots, x_n)$, если $m \geq n$ и для $\forall k (0 \leq k \leq n)$ верно $x_k \leq y_k$. Для обозначения этой ситуации будем писать $\mathbf{x} \leq \mathbf{y}$.

Приведем определения, отражающие элементарные желаемые свойства индекса научной эффективности, выражаемого неотрицательным целым числом.

Определение 1. Индекс научной эффективности — это функция f, отображающая множество X на множество неотрицательных целых чисел $N_0 (X \to N_0)$ и удовлетворяющая следующим двум условиям:

а) если $\mathbf{x}= (0, \ldots, 0)$ или $\mathbf{x}$ — пустой вектор, то $f(\mathbf{x})= 0$;

б) монотонности: если $\mathbf{x} \leq \mathbf{y}$, то $f(\mathbf{x}) \leq f(\mathbf{y})$.

Определение 2. Индекс Хирша h — это функция $h : X \to N$, которая присваивает вектору $\mathbf{x}= (x_1, \ldots, x_n)$ значение

$$h(\mathbf{x}) = \max\{k : x_m \geq k\}, \text{где} m \leq k.$$

Теперь сформулируем аксиомы, касающиеся добавления единичной публикации. Увеличение индекса происходит только за счет публикации с количеством цитирований, большим, чем индекс.

Аксиома A1. Если вектор $\mathbf{y}$ размерности $(n + 1)$ получен из вектора $\mathbf{x}$ размерности n путем добавлением одной публикации с $f(\mathbf{x})$ цитированиями, то $f(\mathbf{y}) \leq f(\mathbf{x})$.

Аксиома A2. Если вектор $\mathbf{y}$ размерности $(n+1)$ получен из вектора $\mathbf{x}$ размерности n добавлением одной публикации с $(f(\mathbf{x})+1)$ цитированиями, то $f(\mathbf{y}) > f(\mathbf{x})$.

Рассмотрим аксиомы, отражающие добавление цитирований к старым работам. Минимальные изменения не должны приводить к большим изменениям индекса.

Аксиома B. Если вектор $\mathbf{y}$ размерности n получен из вектора $\mathbf{x}$ размерности n путем добавления количества цитирований единичной публикации, то $f(\mathbf{y}) \leq f(\mathbf{x})+1$.

Аксиома C. Если вектор $\mathbf{y}$ размерности n получен из вектора $\mathbf{x}$ размерности n путем добавления не более одного цитирования к каждой публикации, то $f(\mathbf{y}) \leq f(\mathbf{x})+1$.

Финальная аксиома, отражающая изменения и в количестве публикаций, и в количестве цитирований:

Аксиома D. Если вектор $\mathbf{y}$ размерности $(n+1)$ получен из вектора $\mathbf{x}$ размерности n путем добавления одной публикации с $f(\mathbf{x})$ цитированиями, а затем увеличением количества цитирований каждой публикации по меньшей мере на единицу, то $f(\mathbf{y}) > f(\mathbf{x})$.

Заметим, что приведенные выше аксиомы не являются независимыми, так из аксиомы A2 следует аксиома D. Показано, что если индекс удовлетворяет аксиомам A1 и A2, то он не должен удовлетворять аксиоме B, и что ни один индекс научной эффективности не может удовлетворять всем аксиомам. Доказана теорема: индекс научной эффективности $f : X \to \mathbb{N}$ удовлетворяет аксиомам A1, B и D тогда и только тогда, когда это h-индекс.

В ряде работ продолжено развитие аксиоматики. Так, в работах [15, 16] предложена аксиоматика для индекса научной эффективности в случае, если индекс задает отображение множества X на множество неотрицательных вещественных чисел. В работе [17] показано, что h-индекс может быть аксиоматизирован с учетом только одного типа изменений, а именно с учетом изменения количества публикаций.

3. Теоретические модели. Теоретическая интерпретация свойств h-индекса на основе математических моделей рассматривается рядом авторов. Здесь основной задачей является статистическое обоснование концепции, заложенной в определении индекса.

3.1. *Модель Хирша.* В своей основополагающей работе [1] Х. Хирш предложил математическую модель h-индекса, в которой связал h-индекс с общим количеством цитирований ($N_{c,tot}$). Публикации, находящиеся в h-ядре, имеют число цитирований не менее h^2, поэтому целесообразно определить коэффициент пропорциональности a, такой что

$$N_{c,tot} = \alpha h^2,$$

а значит

$$h = \sqrt{\frac{N_{c,tot}}{\alpha}}.$$

Этот коэффициент определяет долю неучтенных цитирований и зависит от индивидуального распределения. Эмпирически определено, что значение константы a находится в диапазоне от 3 до 5.

В рамках простейшей линейной модели предполагается, что количество новых цитирований публикаций автора за год есть некоторая константа c, а количество публикаций за год — p. Тогда количество цитирований за $(n+1)$ год выражается формулой

$$N_{c,tot} = \sum_{j=1}^{n} pcj = \frac{pcn(n+1)}{2}.$$

Предположим, что все публикации, вплоть до года y, вносят вклад в вычисление h, тогда верны два равенства: $py = h, (n-y)c = h$. Отсюда

$$h = \frac{c}{1+\frac{c}{p}} n.$$

В результате для достаточно большого n имеем

$$N_{c,tot} \approx \frac{\left(1+\frac{c}{p}\right)^2}{\frac{2c}{p}} h^2.$$

Отсюда следует, что коэффициент a зависит от количества публикаций и цитирований на публикацию, прирастающих за год. Для модели, в которой ученый публикует работы эквивалентного качества и в постоянном темпе, справедливо выражение $h \sim mn$.

Х. Хирш предположил, что отношение h к n может использоваться для сравнения ученых с различной продолжительностью карьеры. При этом первая публикация не всегда может быть подходящей точкой отсчета, так как может пройти время до того, как ученый начнет получать устойчивые результаты. Кроме того, коэффициент не подходит для оценки ученых, не поддерживающих уровень продуктивности в течение всей карьеры.

Если в представленной линейной модели рассматривать публикации по убыванию количества цитирований, X_r^* является линейной функцией от r (X_0^* — число цитирований, полученное наиболее цитируемой публикацией). В этом случае распределение цитирований можно выразить формулой

$$X_r^* = X_0^* - \left(\frac{X_0^*}{h} - 1\right) r.$$

3.2. *Модель Schubert — Glänzel.* В работах [13, 18] исследована зависимость h-индекса от параметров распределения и размеров выборки с использованием теории экстремальных значений Э. Гумбеля [19].

Пусть X — случайная переменная, в нашем случае это темп цитирования публикации. Рассмотрим плотность распределения $X : p_k = P(X = k)$ и функцию распределения $X : F(k) = P(X < k)$.

Определим

$$G_k = G(k) := 1 - F(k) = P(X \geq k).$$

Предположим, имеется выборка из n элементов ($\{X_i\}, i = 1, \dots, n$), где все компоненты независимы и имеют функцию распределения F. Для выборки размером n r-е характеристическое экстремальное значение Гумбеля (u_r) определяется следующим образом:

$$u_r := G^{-1}(r/n) = \max\{k : G_k \geq r/n\}.$$

Ранговая статистика $R(r) = X_r^*$ ($X_1^* \geq X_2^* \geq \ldots \geq X_n^*$ — элементы выборки, упорядоченные по убыванию) может рассматриваться как статистическая оценка r-го наибольшего значения u_r [Glänzel, 2008].

Согласно работе [13], теоретический h-индек C можно определить следующим образом:

$$h := \max\{r : u_r \geq r\} = \max\{r : \max\{k : G_k \geq r/n\} \geq r\}$$

в предположении, что $n > 0,\ X_1 \geq 1$.

Если существует индекс r, такой, что $u_r = r$, то $h := r$ и, следовательно, $h := u_h$.

Рассмотрим важный случай, а именно дискретные распределения Парето с конечным математическим ожиданием. К этой категории принадлежит большинство распределений, используемых для моделирования публикационной активности и процесса цитирования.

Будем считать, что распределение случайной величины является распределением Парето, если оно асимптотически подчиняется закону Ципфа [20], т. е. если k стремится к бесконечности, то $G_k k^{-\alpha}$ — константа. Асимптотически случайные переменные, имеющие распределение Парето (второго типа), удовлетворяют этому условию, так как

$$p_k = P(X = k) \approx d(N + k)^{-(\alpha+1)},$$

если $k \gg 1; \alpha > 1$; N и d — положительные константы. Далее рассматриваются только такие распределения. Для $k \gg N$ выполняются соотношения

$$p_k = P(X = k) \approx dk^{-(\alpha+1)}, \quad G_k = P(X \geq k) \approx d_1 k^{-\alpha},$$

где d_1 — положительная константа. Значит, имеем ожидаемое значение

$$E(X) = \sum_{k=0}^{\infty} kp_k = \sum_{k=0}^{\infty} G_k < \infty, \text{ если } \alpha > 1.$$

Произведя элементарные манипуляции с функцией распределения, можем получить следующую аппроксимацию из определения r-х характеристических экстремумов Гумбеля

$$u_r \approx c_1 \left(\frac{n}{r}\right)^{\frac{1}{\alpha}}, \tag{1}$$

где c_1 — положительная константа. Применяя эти приближения для случая h-индекса, получаем следующее свойство:

$$h = u_h \approx c_1 \left(\frac{n}{h}\right)^{\frac{1}{\alpha}}, \text{ если } n \gg 1. \tag{2}$$

Следовательно,

$$h \approx c_2 n^{1/(\alpha+1)}, \quad \text{если} \quad n \gg 1, \tag{3}$$

где $c_2 = c_1^{\overline{\alpha+1}}$ — положительная константа. Иными словами, h-индекс приблизительно пропорционален корню $(\alpha + 1)$-й степени из числа публикаций.

В случае журналов представляет интерес определение h-индекса через такие параметры распределения цитирований, как математическое ожидание — среднее значение числа цитирований на одну публикацию (обозначим его CPP) и „размер выборки“. В случае распределения Парето (второго типа) с двумя параметрами N, α ожидаемое значение выражается равенством

$$CPP = \frac{N}{\alpha - 1} \tag{4}$$

(константа в выражении (**??**) $c_1 = N$). Тогда, согласно (**??**) и (**??**),

$$h = u_h \approx c_1 \left(\frac{n}{h}\right)^{\frac{1}{\alpha}} \approx N^{\alpha/(\alpha+1)} n^{1/(\alpha+1)}. \tag{5}$$

В случае $\alpha = 2$, что соответствует закону Лотки с экспонентой 3, согласующемуся с предположениями, принятыми в библиометрии, выражение (**??**) можно записать как $CPP = N$. Тогда, преобразуя (**??**), получаем

$$h = cn^{1/3} CPP^{2/3}, \tag{6}$$

где c — положительное число порядка 1.

Связь между h, n и CPP изучалась на практике. Результаты приведены в работе [18]. Данные взяты из *WoS*. За 2001 г. выбраны 6406 журналов, а за 2002 г. — 6481. Рассмотрено трехгодовое окно цитирования (год публикации и два предшествующих), чтобы иметь возможность вычислять как h-индекс, так и импакт-фактор *IF* (в этом случае *CPP* = *IF*). Во внимание приняты четыре типа документов: статьи (англ. *articles*), письма (англ. *letters)*, заметки (англ. *notes*) и обзоры (англ. *reviews*). Показано, что не только h имеет линейную корреляцию с произведением $n^{1/3} IF^{2/3}$, но и что константа c независима от научной области и года рассмотрения и колеблется вблизи значения 0,75.

Таким образом, теоретическая связь между h, n и *CPP*, выведенная из модели Парето для распределения цитирований публикаций журналов и представленная в выражении (**??**), полностью соответствует эмпирическим данным о цитировании за год и окне цитирования, равном трем годам.

Заметим, что единственный параметр c практически не зависит от научной области. Отсюда следует, что индекс h не имеет особой зависимости от научной области, кроме хорошо известной зависимости от характерных для области количества публикаций и скорости цитирования. В этом смысле выражение (**??**) обеспечивает некоторую „трансформацию подобия“ h-индекса между различными областями.

3.3. *Модель Egghe — Rousseau.* В работе [21] h-индекс определяется в рамках обобщенной концепции *Information Production Processes* (*IPP*s, процессы производства информации) в терминах „источник — объект“ (англ. *source — item*). Примером пары „источник — объект“ могут служить пары „журнал — статья“ или „статья — цитирование“.

Рассмотрим *IPP,* состоящий из источников и объектов. Пусть $R(r)$ — функция ранжирования этой системы: если источники упорядочены в убывающем порядке количества объектов, то в дискретном случае $R(r)$ — количество объектов, „производимых“ источником ранга r (будем считать эту функцию непрерывной). Переформулируем определение h-индекса для данной концепции, используя T вместо N_p из первоначального определения. В непрерывном случае функция R, определенная на отрезке $[0, T]$ и задающая плотность объектов, полагается строго положительной, строго убывающей и непрерывной. Тогда h-индекс — это r, такое, что $R(r) = r$.

В работе [21] показано, что каждый *IPP* имеет единственный h-индекс.

Теперь в системе, в которой выполняется степенной закон Лотки (о зависимости между количеством авторов и их научной производительностью) [22], рассмотрим функцию зависимости

$$f : [1, \infty[\to]0, C]$$

вида:

$$f(j) = \frac{C}{j^{\alpha}},$$

где $C > 0$, $\alpha > 1$ — экспонента Лотки. В дискретном варианте функция определяет количество источников продуктивности j. В непрерывном варианте функция интерпретируется как плотность распределения. Показано, что в рамках закона Лотки для заданного количества источников T верно

$$h = T^{\frac{1}{\alpha}}.$$

3.4. *Сравнение моделей.* В работе [23] проведен сравнительный анализ трех представленных выше моделей на материале *WoS* и *Essential Scientific Indicators* (*ESI*, *Thomson Reuters*) применительно к журналам и организациям. Для единообразия переформулируем представление h-индекса в соответствии с математическими моделями. Пусть C — общее количество цитирований, P — количество публикаций.

В модели Хирша $h = \left(\frac{C}{a}\right)^{1/2}$, где $3 < a < 5$ — константа. В модели Egghe — Rousseau $h = P^{1/\alpha}$, где α — константа Лотки. В модели Schubert — Glänzel $h = cP^{1/3}(CPP)^{2/3}$, где *CPP*$= C/P$ — среднее количество цитирований, которое для журналов ассоциируется с импакт-фактором *IF*, c — константа.

Проведены исследования следующих оценок h-индекса, соответствующих моделям, при $\alpha = 2$, $a = 5$:

$$h_c \sim \sqrt{\frac{C}{5}}, \quad h_p \sim \sqrt{P}, \quad h_{pc} \sim cP^{\frac{1}{3}}\left(\frac{C}{P}\right)^{\frac{2}{3}}.$$

Данные об эмпирическом h-индексе брались из БД *WoS* в рамках временного окна, равного десяти годам, а данные о P, C, *CPP* — из БД *ESI*. При вычислении h_{pc} константа c устанавливалась равной c =0,9 для журналов и $c = 1$ для организаций.

Результаты анализа показали, что оценка в модели *Schubert — Glänzel* ближе к реальным значениям h-индекса как для журналов, так и для организаций. Поскольку при вычислении h-индекса учитываются и цитирования, и публикации, эта модель наиболее близка к реальности. В большинстве случаев сравнительная характеристика для журналов и организаций имеет вид

$$h_p < h \sim h_{pc} < h_c.$$

4. *h*-последовательности и *h*-матрицы. Рассмотрим работы, касающиеся расширения возможностей h-индекса и восполнения отсутствующих деталей истории цитирования.

4.1. *h-последовательность Liang.* Определение h-последовательности (англ. *h-sequence*) как средства для изучения механизма изменения h-индекса во времени и сравнения ученых с различными периодами карьеры введено в работе [24]. h-последовательность — это последовательность h-индексов, вычисленных за увеличивающиеся промежутки времени в обратном порядке по годам, начиная с текущего года.

Например, фиксируем 2004-й год, вычисляем h-индекс, получаем h_1 — первый элемент h-последовательности. Вычисляем h-индекс за период 2004–2003 гг., получаем h_2 — второй элемент h-последовательности. Далее вычисляем h-индекс за 2004–2002 гг., получаем h_3 — третий элемент h-последовательности. Выполняем аналогичные вычисления до некоторого фиксированного года. Например, если рассматривается период 2004–1976 гг., получится последовательность h_1, h_2, ..., h_{29}.

Для ранжирования ученых по этому принципу построим h-матрицу, столбцы которой соответствуют ученым, а строки — элементу h-последовательности: h_1, h_2, ..., h_n.

h-последовательность выявляет механизм увеличения h-индекса, а h—матрица позволяет сравнивать результативность ученых с различной научной карьерой, если взять за начало год, до которого все ученые уже опубликовали первую статью. Можно построить новую матрицу для всех ученых за первые n лет. Возможен другой подход: для каждого ученого выбрать n самых продуктивных лет, а затем провести сравнение. В этом случае следует выбрать окно цитирования, например, подсчитывать цитирования через m лет после публикации.

4.2. *h-последовательность Egghe.* В работе [25] предложено начинать строить h-последовательности с начала карьеры ученого. Пусть период карьеры описывается временными отрезками $t = 1, 2, \ldots, t_m$, где $t = 1$ соответствует первому году карьеры (первой публикации), а t_m — последнему (ограничившись, например, текущим годом). h-последовательность строится следующим образом. Для $t = 1$ рассматриваются только публикации и цитирования текущего года. Вычисляем h-индекс h_1. Для $t = 1$ и $t = 2$ вместе вычисляем h-индекс h_2, рассматривая все цитирования и публикации за этот период и т. д. Последовательность $h_1, h_2, \ldots, h_{t_m}$ дает динамическое представление о карьере ученого и может использоваться для сравнения результативности ученых. Заметим, что h-индексы для h-последовательности *Liang* (обозначим как $h_1^*, h_2^*, \ldots, h_{t_m}$) автоматически получаются на основе данных из *WoS*, в то время как h-индексы для h-последовательности *Egghe* необходимо вычислять вручную.

В работе [25] рассматривается непрерывный временной интервал $t \in R^+$ и вычисляются указанные два типа последовательностей в рамках среды, где выполняется закон Лотки, связывающий источники и объекты (здесь публикации и цитирования):

$$f(j) = \frac{C}{j^{\alpha}},$$

где $C > 0$; $\alpha > 1$; $f\ (j)$ — плотность публикаций с плотностью j для цитирований. Предполагается, что α — константа для каждого периода времени. Поскольку временной интервал является непрерывным, вместо $h_1, h_2, \ldots, h_{t_m}$ рассматривается $h(t)$, вместо $h_1^*, h_2^*, \ldots, h_{t_m}^*$ — $h(t)^*$.

Показано, что в общем случае графики функции различаются, т. е. последовательности ведут себя неодинаково. Это подтверждено эмпирически.

В отличие от единичного h-индекса, h-последовательности иллюстрируют эволюцию карьеры ученого.

4.3. *h-последовательность Randic.* Иной подход к расширению информации, используемой при вычислении h, предложен в работе [26]. Рассмотрим ранжированный по количеству цитирований список публикаций автора. При вычислении h ранги нумеруются от 1 до n (количество работ) и вычисляется

$$h = \max\{r : X_r^* \geq r\}.$$

Обозначим $H_0 = h$. Для построения H_1 удваиваем ранги и вновь вычисляем h. Вычисления продолжаем до тех пор, пока есть работы, цитирование которых превосходит ранг. Получаем последовательность $\{^0H,\ ^1H, \ldots\}$, из которой следует H-последовательность $\{^0H,\ ^0H +^1 H,\ ^0H +^1 H +^2 H, \ldots\}$. Последняя сумма и есть H-индекс H. Теперь вычисляем коэффициент $Q = H/h$, который рассматривается как мера потенциала ученого, построенная на основе истории цитирования.

В работе [27] данный подход изучается в рамках системы, в которой выполняется закон Лотки. Определяются формулы для H-индекса и H-последовательности, а также необходимые и достаточные условия, при которых $H_1 > H_2$ в случае равенства h-индексов двух ученых ($h_1 = h_2$). Новый индекс и последовательности дают дополнительную характеристику манере цитирования и позволяют ранжировать ученых, имеющих равные h-индексы.

Заключение. Индекс Хирша и другие подобные индексы, основанные на статистическом анализе библиометрических данных, значительно расширяют горизонты наукометрии. На сайте „*h-indexand Variants*“ [http://sci2s.ugr.es/hindex/] представлено более сорока модификаций h-индекса. Отметим, что появление любой новой модификации еще раз подтверждает факт, что с помощью одного показателя невозможно оценить уровень научной продуктивности.

Список литературы

1. Hirsch J. E. An index to quantify an individual's scientific research output // Proc. of the National Acad. Sci. USA. 2005. V. 102, N 46. P. 16569–16572.

2. Hirsch J. E. An index to quantify an individual's scientific research. [Electron. resource]. http://xxx.arxiv.org/abs/physics/0508025.

3. Glänzel W. On the opportunities and limitations of the H-index // Sci. Focus. 2006. V. 1, N 1. P. 10–11.

4. Bornmann L., Daniel HD. What do we know about the h index? // J. Amer. Soc. Inform. Sci. Tech. 2007. V. 58, iss. 9. P. 1381–1385.

5. Bornmann L., Daniel HD. Does the h-index for ranking of scientists really work? // Scientometrics. 2005. V. 65, iss. 3. P. 391–392.

6. Cronin B., Meho L. Using the h-index to rank influential information scientists // J. Amer. Soc. Inform. Sci. Tech. 2006. V. 57, iss. 9. P. 1275–1278.

7. Van Raan A. F. J. Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups // Scientometrics. 2006. V. 67, iss. 3. P. 491–502.

8. Kelly C. D., Jennions, M. D. The h-index and career assessment by numbers. // Trends in Ecology & Evolution. 2006. V. 21, N 4. P. 167–170.

9. Sidiropoulos A., Katsaros D., Manolopoulos Y. Generalized Hirsch h-index for disclosing latent facts in citation networks // Scientometrics. 2007. V. 72, iss. 2. P. 253–280.

10. Lehnmann S., Jackson A. D., Lautrup B. E. Measures and mismeasures of scientific quality. [Electron. resource]. http://arxiv.org/abs/physics/0512238.

11. Braun T., Glänzel W., Schubert A. A. Hirsch-type index for journals // Scientometrics. 2006. V. 69, iss. 1. P. 169-173.

12. Braun T., Glänzel W., Schubert A. A. Hirsch-type index for journals //Te Scientist. 2005. V. 19, iss. 22, P. 8. [Electron. resource]. http://www.the-scientist.com/?articles.view/articleNo/16863/title/A-Hirsch-type-index-for-journals/

13. Glänzel W. On the h-index: A mathematical approach to a new measure of publication activity and citation impact // Scientometrics. V. 67, iss. 2. 2006. P. 315–321.

14. Woeginger G. J. An axiomatic characterization of the Hirsch-index // Mathematical Social Sci. 2008. V. 56. P. 224–232.

15. Quesada A. Monotonicity and the Hirsch index // J. Informetrics. 2009. V. 3, iss. 2. P. 158–160.

16. Quesada A. More axiomatics for the Hirsch index // Scientometrics. 2010. V. 82, iss. 2. P. 413–418.

17. Miroiu A. Axiomatizing of the Hirsch index: quantity and quality disjoined // J. Informetrics. 2013. V. 7, iss.1. P. 10–15.

18. Schubert A., Glänzel W. A systematic analysis of Hirsch-type indices for journals // J. Informetrics. 2007. V. 1, iss. 3. P. 179–184.

19. Гумбель Э. Статистика экстремальных значений. М.: Мир, 1965.

20. Zipf G. K. Human behavior and the principle of least effort, Addison-Wesley Press, 1949.

21. Egghe L., Rousseau R. An informetric model for the Hirsch index // Scientometrics. 2006. V. 69, iss. 1. P. 121–129.

22. Lotka A. J. The frequency distribution of scientific productivity. J. Washington Academy of Sciences. 1926. V. 16. N. 12. P. 317–324.

23. Ye F. Y. An investigation on mathematical models of the h-index // Scientometrics. 2009. V. 81, iss. 2. P. 493–498.

24. Liang L. h-index sequence and h-index matrix: Constructions and applications // Scientometrics. 2006. V. 69, iss. 1. P. 153–159.

25. Egghe L. Mathematical study of h-index sequences // Inform. Proc. Management. 2009. V. 45, iss. 2. P. 288–297.

26. Randič M. Citations versus limitations of citations: beyond Hirsch index // Scientometrics. 2009. V. 80, iss. 3. P. 809–818.

27. Egghe L. Mathematical results on the Randiж's H-index and H-sequence // Research evaluation. 2010. V. 19, iss. 3. P. 203–207.

Бредихин Сергей Всеволодович — канд. техн. наук,
зав. лабораторией Института вычислительной
математики и математической геофизики СО РАН;
e-mail: bred@nsc.ru;
Щербакова Наталья Григорьевна — ст. науч. сотр.
Института вычислительной математики
и математической геофизики СО РАН;
e-mail: nata@nsc.ru

Дата поступления — 18.04.2014