

АЛЬТЕРНАТИВНАЯ ДРОБНАЯ МОДИФИКАЦИЯ ИНДЕКСА ХИРША, УЧИТЫВАЮЩАЯ КОЛИЧЕСТВО АВТОРОВ ЦИТИРУЕМЫХ СТАТЕЙ

Марвин С. В.¹

*(ФГАОУ ВПО «Уральский Федеральный университет
им. первого Президента России Б.Н. Ельцина»,
Екатеринбург)*

Изучены недостатки индекса Хирша и других сходных наукометрических показателей, при вычислении которых не учитывается количество соавторов или учитывается с некоторыми логическими ошибками. Предложен дробный аналог индекса Хирша, который учитывает число соавторов, а также с достаточной полнотой и точностью характеризует индивидуальный научный вклад ученого, работающего в областях науки с традиционно невысокой цитируемостью.

Ключевые слова: наукометрия, цитируемость, индекс Хирша, соавторство, ядро Хирша.

1. Индекс Хирша и существующие его модификации, учитывающие соавторство

В настоящее время для оценки результативности публикационной деятельности ученых используются различные наукометрические показатели [7], основным из которых является индекс Хирша или h -индекс [9]. Коллективное значение h -индекса используется для оценки публикационной деятельности отдельных подразделений и организаций в целом (но все же по отдельным областям науки: общеизвестно, что сравнивать индексы Хирша ученых, задействованных в разных отраслях

¹ Сергей Владимирович Марвин, кандидат физико-математических наук, доцент (s.v.marvin@yandex.ru).

науки, бессмысленно [3–5, 7] в силу сильной специфики каждой отдельной отрасли). В отличие от простого количества цитирований (без учета количества публикаций) или простого количества публикаций (без учета цитируемости) индекс Хирша характеризует и публикационную активность, и цитируемость (это является основным достоинством h -индекса). Для высокого значения индекса Хирша нужно, чтобы у автора было не просто много публикаций и не просто много цитирований, а много публикаций, каждую из которых много раз процитировали. Точная формулировка правила, по которому вычисляется индекс Хирша, такова: индекс Хирша данного автора равен h , если среди всех его работ есть h публикаций, на каждую из которых сослались не менее h раз, и в то же время на каждую из остальных его публикаций сослались не более h раз. Те h публикаций, каждую из которых процитировали не менее h раз, образуют, по терминологии наукометрии, ядро Хирша (также будем называть его h -ядром). Совокупность остальных публикаций автора называется периферией, или «хвостом» [7].

Для большей объективности при расчете индекса Хирша не следует учитывать самоцитирования (в библиографической базе elibrary.ru для этого вычисляется отдельный показатель – индекс Хирша без самоцитирований).

Один из основных недостатков индекса Хирша заключается в том, что при его расчете не учитывается количество авторов публикаций [4, 5, 7]. Индекс Хирша призван характеризовать личный вклад автора в науку, однако он считается одинаково как для авторов, не имеющих соавторов, так и для авторов, имеющих многочисленных соавторов.

Самый простой из всех ранее предлагавшихся способов учета количества соавторов заключается в делении индекса Хирша на среднее число авторов публикаций, образующих h -ядро: так вычисляется индекс, называющийся в наукометрии *individual h -index* и обозначаемый как h_I [6]. Однако сам принцип формирования h -ядра, изначально учитывающий только число цитирований, но не соавторов, говорит не в пользу достоверности индекса h_I .

Самим Хиршем была предложена модификация индекса, при вычислении которой цитирования каждой статьи нужно

распределять между соавторами в соответствии с текущими значениями их h -индексов; вычисляемый по такому принципу индекс обозначается $hbar$ [10]. Однако, как уже было замечено, вычисление такого индекса сопряжено с большими техническими трудностями [4]: необходимо по библиографическим базам для каждой научной работы определять $hbar$ авторов на момент публикации. Кроме того, представляется очень спорным утверждение, что вклад автора в конкретную публикацию соответствует значению его $hbar$.

Также следует заметить, что вклад ученого в конкретную публикацию не всегда можно определить по порядковому номеру его фамилии в списке авторов научной работы. Разные журналы имеют разные традиции, касающиеся расположения фамилий авторов в заголовке статьи. Например, в «Журнале вычислительной математики и математической физики» фамилии авторов принято располагать в алфавитном порядке, и никак иначе. А в журнале «Дефектоскопия» порядок перечисления авторов определяется самими авторами. Встречается также точка зрения, что первый из указанных авторов должен быть первым по значимости, а остальные равнозначны между собой, и фамилии авторов, начиная со второго, должны располагаться в алфавитном порядке [2]. В электронных библиографических базах учесть все эти нюансы невозможно. Поэтому для выработки общих подходов к анализу цитируемости наилучшим средством представляется простое деление числа цитирований публикации на число ее авторов [4, 5].

Если при вычислении индекса Хирша для каждой работы вместо обычного числа цитирований использовать число цитирований, деленное на число авторов, произойдет резкое понижение индекса Хирша ученых, публикующихся с многочисленными соавторами, что, несомненно, сделает этот индекс более достоверной оценкой научной работы. Заметим, что в наукометрии так вычисляемый индекс уже рассматривался [7]. Для такого индекса есть специальное название: individual h -index (PoP variation); далее будем обозначать его h_{IP} . В какой-то степени этот показатель анализировался: были рассмотрены конкретные примеры, как h_{IP} меняет картину научных достижений некоторых высокоцитируемых ученых [4]. Однако и у этого

индекса есть недостатки, которые можно не увидеть при неполном изучении темы. Изложим их суть.

Индекс Хирша, даже не учитывающий число соавторов, малочувствителен к увеличению числа цитирований на 1. Поясним это. Предположим, мы рассматриваем список публикаций конкретного ученого, ранжированный в порядке убывания числа цитирований, и определяем индекс Хирша. Если на h -м месте списка располагается публикация, которую h раз процитировали, то индекс Хирша ученого равен h . Однако если число цитирований h -й публикации увеличить до $h + 1$, а на $(h + 1)$ -м месте все еще будет публикация с числом цитирований, не большим h , то индекс Хирша так и останется равным h (при этом h -я публикация, в зависимости от количества ссылок на публикации с меньшими номерами, может поменяться с ними местами в списке, т.е. переместиться вглубь ядра). Если число ссылок на $(h + 1)$ -ю публикацию увеличится до $h + 1$, но число ссылок на h -ю публикацию так и останется равным h , то произойдет перестановка статей в нашем ранжированном списке, а индекс Хирша так и останется равным h . Только если число цитирований и у h -й, и у $(h + 1)$ -й публикации повысится до $h + 1$, причем при условии, что на все остальные публикации в ядре Хирша будет не менее $h + 1$ ссылок, индекс Хирша повысится до $h + 1$.

Если научная работа написана с соавторами, то число ее цитирований, разделенное на число авторов, будет увеличиваться даже не на единицу, а на дробные доли, вследствие чего динамика h_{IP} в некоторых областях науки будет очень медленной. Например, для физико-математических и технических наук совсем не типичны частые и многочисленные цитирования, в отличие от биологии и медицины [3, 4, 7]. Для материаловедения обычным является количество авторов, не меньшее пяти [2]. Вполне ожидаемо, что в математике, физике и технических науках при сложившихся традициях цитирования индекс h_{IP} у подавляющего большинства ученых вообще не будет меняться в течение многих лет или даже десятилетий, несмотря на ненулевую публикационную активность и ненулевую цитируемость этих ученых. Только по истечении чрезвычайно длительного промежутка времени h_{IP} будет скачком изменяться с h до $h + 1$, так как этот показатель является целочисленной величиной и

между h и $h + 1$ нет промежуточных вариантов. Такой индекс, конечно, не может считаться адекватной характеристикой публикационной активности и цитируемости автора (особенно если цитируемость дробная).

Характерно, что интересные примеры применения h_{IP} с подробными расчетами [4] относятся к ученым, обычные индексы Хирша которых составляют 25, 47 и 23, а h_{IP} равны, соответственно, 10, 15 и 18. Фамилии и даже области научных интересов этих ученых не разглашаются, но имеет смысл заметить, что в математике, физике и технических науках такие значения индекса Хирша соответствуют уровню член-корреспондентов и академиков РАН.

Следует ожидать, что в областях науки, в которых подавляющее большинство ученых имеет индекс Хирша между 0 и 9, h_{IP} будет принимать значения 0, 1, 2 и 3 без каких-либо промежуточных (дробных) вариантов. Это приведет к недопустимому уравниванию научных достижений, разных по существу.

Указанные недостатки приводят к выводу, что для корректировки индекса Хирша в соответствии с числом соавторов необходимо предложить некоторую дробную модификацию h_{IP} . При этом хотелось бы сохранить основную идею, связанную с индексом Хирша: для больших значений индекса должно быть много публикаций, на каждую из которых много раз сослались. Эту идею лучше всего разобрать на конкретном примере с иллюстрацией.

На рис. 1 изображен график цитируемости некоторого предполагаемого автора. Сплошная линия показывает зависимость числа цитирований публикации от ее номера в списке всех работ автора (если список составлен в порядке убывания числа цитирований). Номер публикации n откладывается по горизонтальной оси x , а число ее цитирований c_n – по вертикальной оси y . Отдельные точки, полученные по такому принципу, соединены ломаной. При этом ломаная дополнена горизонтальным отрезком слева: высота этого отрезка равна цитируемости первой статьи. То есть можно считать, что введена «мнимая» нулевая публикация, число цитирований которой равно числу цитирований первой публикации. Если у автора N работ, на последнюю из которых в нашем ранжированном спис-

ке есть ссылки, то ломаная справа дополняется еще одним наклонным отрезком с концами (N, c_N) и $(N + 1, 0)$. То есть вводится «мнимая» $(N + 1)$ -я публикация с количеством ссылок, равным нулю. Практический смысл «мнимых» публикаций будет разъяснен ниже. Пока только заметим, что при оценке цитируемости автора наличие $(N + 1)$ -й публикации, которую не цитируют, равносильно ее отсутствию.

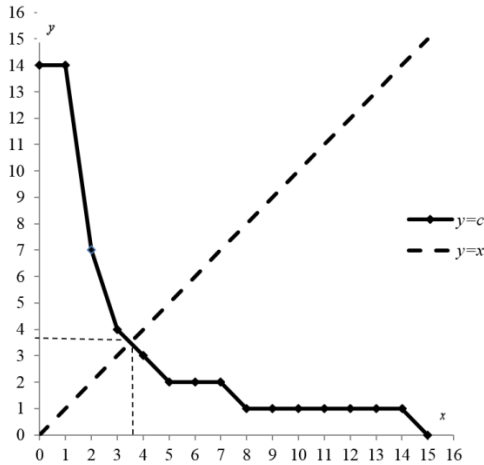


Рис 1. Графическое определение индекса Хирша

Ломаная также может завершаться и горизонтальными звеньями, лежащими на оси x , если в конце ранжированного списка научных работ автора есть две или большее количество публикаций с нулевой цитируемостью. Но эти звенья, лежащие на оси x , не играют роли при определении индекса Хирша.

Для поиска публикации, номер которой в ранжированном списке равен индексу Хирша, проводится прямая $y = x$ (на рис. 1 она изображена крупным пунктиром). Точка пересечения прямой и ломаной определяет границу между ядром (узлы ломаной над прямой) и периферией (узлы ломаной под прямой). Индекс Хирша равен целой части абсциссы этой точки (равно как и ординаты: точка располагается на прямой $y = x$). То есть при определении h -индекса играет роль только граница между ядром и периферией. На индекс Хирша не повлияет ни сколь

угодно большая цитируемость статей, располагающихся в начале ядра, ни маленькое увеличение числа цитирований на каждую отдельную статью в периферии (даже если суммарное число цитирований сильно возрастет). В ситуации, проиллюстрированной на рис. 1, абсцисса точки пересечения ломаной и прямой приблизительно равна 3,7; индекс Хирша $h = 3$.

Ранее предлагались дробные модификации индекса Хирша, призванные скомпенсировать медленную и скачкообразную динамику обычного h -индекса в математике: Sh и h_{rat} [6, 8]. Однако эти модификации не учитывают количество авторов и отступают от основополагающей идеи h -индекса: на Sh и h_{rat} сильно сказывается динамика всего h -ядра, а динамика границы между ядром и периферией оказывается недостаточно учтенной.

Предлагаемая в данной статье модификация индекса Хирша основана на следующих предположениях. Во-первых, для наиболее объективной характеристики индивидуального вклада ученого в науку необходимо при расчете индекса исключить самоцитирование. Во-вторых, для построения ломаной, аналогичной изображенной на рис. 1, необходимо для каждой конкретной публикации в качестве c_n использовать не обычное число цитирований, а число цитирований, разделенное на количество авторов. Ранее уже было предложено название этой величине – долевое число цитирований [4]; будем пользоваться предложенным термином. В-третьих, для нового значения индекса нужно брать не целую часть абсциссы, а саму абсциссу точки пересечения прямой и ломаной на графике цитируемости. Указанную альтернативу индексу Хирша будем называть модифицированным индексом Хирша и обозначать h_{mod} .

2. Качественный анализ предложенной дробной модификации индекса Хирша и формула для расчета

Заметим, что даже если автор публикаций никогда не имел соавторов и удельная цитируемость каждой его статьи совпадает с обычной, то h_{mod} все равно может не совпадать с h , так как h является целой частью h_{mod} . То есть и в такой ситуации h_{mod} имеет уточняющий смысл для h .

Заметим также, что при замене цитируемости на долевую цитируемость может произойти перестановка публикаций в списке, используемом для вычисления обычного индекса Хирша. Например, если у ученого первая публикация имеет 7 ссылок, а вторая публикация имеет 4 ссылки, то первая публикация имеет более высокую цитируемость. Однако если у первой публикации 5 авторов (т.е. еще 4, помимо рассматриваемого автора), а у второй число авторов равно 2, то долевая цитируемость первой публикации равна 1,4, а у второй она равна 2. Следовательно, в списке, ранжированном по долевой цитируемости, вторая публикация будет располагаться перед первой.

Также следует заметить, что при введении «мнимых» публикаций ломаная на графике цитируемости соединяется с координатными осями, что гарантирует пересечение прямой $y = x$ с ломаной, даже если первый узел ломаной изначально располагается слишком низко (под прямой $y = x$) или ее последний узел располагается слишком высоко (над прямой $y = x$).

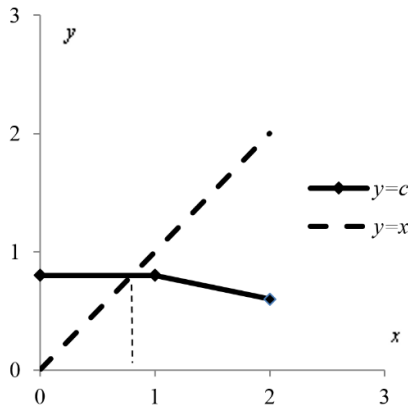


Рис. 2. c_1 не превосходит 1

Разберем подробно, какие значения модифицированного индекса Хирша получаются из графика цитируемости. Если долевое число цитирований у первой статьи в нашем ранжированном списке не превосходит 1, то прямая $y = x$ пересекает первый (горизонтальный) отрезок ломаной, и, как видно из

рис. 2, $h_{mod} = c_1$. С увеличением долевого числа цитирований первой статьи высота горизонтального отрезка будет увеличиваться, точка пересечения будет смещаться вверх и вправо, что приведет к увеличению h_{mod} .

Если же $c_1 > 1$ (рис. 3), то находим такой номер n , для которого $c_n \geq n$ и в то же время $c_{n+1} \leq n + 1$. Далее, пользуясь уравнением прямой по двум точкам [1], приходим к следующему уравнению для h_{mod} :

$$(1) \quad \frac{h_{mod} - c_{n+1}}{c_n - c_{n+1}} = \frac{n + 1 - h_{mod}}{1}.$$

Из уравнения (1) получаем формулу для вычисления модифицированного индекса Хирша:

$$(2) \quad h_{mod} = \frac{c_n \cdot (n + 1) - c_{n+1} \cdot n}{c_n + 1 - c_{n+1}}.$$

Для горизонтальных отрезков ломаной, т.е. при $c_n = c_{n+1}$, значение h_{mod} , получающееся по этой формуле, равно c_{n+1} :

$$h_{mod} = \frac{c_n \cdot (n + 1) - c_{n+1} \cdot n}{c_n + 1 - c_{n+1}} = \frac{c_{n+1} \cdot (n + 1) - c_{n+1} \cdot n}{c_{n+1} + 1 - c_{n+1}} = c_{n+1}.$$

В частности, для ситуации, проиллюстрированной на рис. 2, из формулы (2) получается верное значение h_{mod} :

$$h_{mod} = \frac{c_0 \cdot 1 - c_1 \cdot 0}{c_0 + 1 - c_1} = \frac{c_1}{c_1 + 1 - c_1} = c_1.$$

Следовательно, формула (2) универсальна.

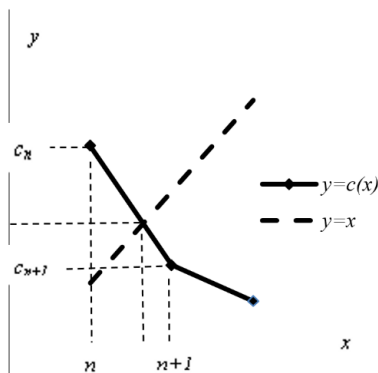


Рис. 3. c_1 больше 1

По рис. 3 легко установить, что если долевое число цитирований у n -й публикации будет расти, то будет меняться наклон $(n + 1)$ -го отрезка ломаной, и точка пересечения этого отрезка с прямой $y=x$ будет перемещаться вверх и вправо, что увеличит h_{mod} . Если будет увеличиваться долевое число цитирований у $(n + 1)$ -й публикации, то, опять же, изменение наклона приведет к перемещению точки пересечения вверх и вправо. Причем если удельное число цитирований превзойдет $n + 1$, то точка пересечения прямой и ломаной переместится на следующее звено ломаной, что еще сильнее увеличит значение h_{mod} . То есть h_{mod} должным образом изменяется при увеличении числа цитирований у n -й и у $(n + 1)$ -й публикации.

Необходимо определить точность, с которой следует вычислять значения c_n и h_{mod} (все нецелые числовые показатели в электронных библиографических базах, естественно, указываются с некоторым округлением). Пусть h_{mod} определяется с точностью до 0,1: такая точность вычисления h_{mod} представляется достаточной для ранжирования ученых в соответствии с их публикационной активностью и цитируемостью. Долевою цитируемость c_n , как промежуточный результат вычислений, следует определять с большей точностью – 0,01.

В качестве примера применения формулы (2) рассчитаем модифицированный индекс Хирша у шести реальных ученых, работающих в одном из разделов технических наук – неразрушающем контроле. Фамилии этих ученых не раскрываются, будем обозначать их первыми шестью буквами латинского алфавита: А, В, С, D, E, F. Необходимые для наших расчетов данные приведены в таблицах 1–6 (следует заметить, что фигурирующие в таблицах показатели цитируемости без учета самоцитирований типичны для указанной области науки даже при более чем двадцатилетнем стаже работы). Как видно из приведенных данных, число авторов у публикаций А–F внутри h -ядра и h_{mod} -ядра меняется в пределах от 1 до 6.

Таблица 1. Публикационная деятельность ученого А

n	1	2	3	4	Статьи с меньшим числом цитирований и меньшим c_n
Количество цитирований	9	7	6	3	...
Количество авторов	2	2	2	3	...
c_n	4,5	3,5	3	1	...

Таблица 2. Публикационная деятельность ученого В

n	1	2	3	4	Статьи с меньшим числом цитирований и меньшим c_n
Количество цитирований	9	5	3	3	...
Количество авторов	4	3	3	3	...
c_n	2,25	1,67	1	1	...

Таблица 3. Публикационная деятельность ученого С

n	1	2	3	4	Статьи с меньшим числом цитирований и меньшим c_n
Количество цитирований	9	3	3	2	...
Количество авторов	2	2	1	1	...
c_n	4,5	1,5	3	2	...

Таблица 4. Публикационная деятельность ученого D

n	1	2	3	4	Статьи с меньшим числом цитирований и меньшим c_n
Количество цитирований	6	6	5	3	...
Количество авторов	5	6	5	4	...
c_n	1,2	1	1	0,75	...

Таблица 5. Публикационная деятельность ученого E

n	1	2	3	4	Статьи с меньшим числом цитирований и меньшим c_n
Количество цитирований	6	5	5	3	...
Количество авторов	6	5	5	4	...
c_n	1	1	1	0,75	...

Таблица 6. Публикационная деятельность ученого F

n	1	2	3	4	5	Статьи с меньшим числом цитирований и меньшим c_n
Количество цитирований	9	6	6	4	4	...
Количество авторов	3	5	5	3	5	...
c_n	3	1,2	1,2	1,33	0,8	...

Заметим, что в таблицах 1–6 публикации упорядочены по количеству цитирований, но не по величине долевой цитируемости. Поэтому в последних строчках таблиц 3 и 6 значения c_n

идут не в порядке убывания, и это нужно иметь в виду при вычислении h_{mod} .

По данным из таблиц 1–6 определяем величины обычных и модифицированных индексов Хирша; результаты объединяем в новую таблицу – таблицу 7. Значения обычного индекса Хирша без учета самоцитирований указаны в eLibrary.ru, но их также можно определить и по представленным в таблицах 1–6 данным (результаты, конечно, совпадают).

Таблица 7. Индекс Хирша: обычный и модифицированный

Ученый	A	B	C	D	E	F
h	3	3	3	3	3	4
h_{mod}	3	1,8	2,5	1,2	1	1,7

По таблице 7 видно, как количество соавторов влияет на модифицированный индекс Хирша. В частности, у F индекс $h = 4$, что превосходит индексы Хирша других ученых из представленной выборки: для A–E индекс $h = 3$. Однако по величине h_{mod} ученый F занимает четвертое место среди всех остальных. Кроме того, резко повлияло на величину h_{mod} большое количество соавторов в публикациях ученых D и E, а также в первой публикации ученого B.

3. Заключение

Предложенный в данной статье модифицированный индекс Хирша полностью устраняет, по всей видимости, самый основной и очевидный недостаток обычного индекса Хирша. При этом вычисление предложенного индекса является технически несложным и может быть выполнено в рамках любой электронной библиографической базы с помощью простых подпрограмм.

Литература

1. БЕКЛЕМИШЕВ Д.В. *Курс аналитической геометрии и линейной алгебры*: Учеб. – 7-е изд. – М.: Высшая школа, 1998. – 320 с.

2. ИВАНОВ К.С. *Индекс Хирша: модернизация необходима (еще раз о количественной оценке работы ученых)* // Троицкий вариант – Наука. Рубрика: Гайд-парк онлайн. – 30.07.2014. – [Электронный ресурс] – URL: <http://trv-science.ru/2014/07/30/indeks-khirsha-modernizaciya-neobkhodima/> (дата обращения: 30.03.2015).
3. МАРШАКОВА-ШЕЙКЕВИЧ И.В. *Роль библиометрии в оценке исследовательской активности науки* // Управление большими системами. – 2013. – № 44. – С. 210–247.
4. МИХАЙЛОВ О.В. *О возможной модификации индексов Хирша и Эгга с учетом соавторства* // Социология науки и технологий. – 2014. – Т. 5, №3. – С. 48–56.
5. ПОЛЯНИН А.Д. *Недостатки индексов цитируемости и Хирша и использование других наукометрических показателей* // Математическое моделирование и численные методы. – 2014. – № 1. – С. 131–144.
6. ШТОВБА С.Д., ШТОВБА Е.В. *Обзор наукометрических показателей для оценки публикационной деятельности ученого* // Управление большими системами. – 2013. – № 44. – С. 262–278.
7. ЦЫГАНОВ А.В. *Краткое описание наукометрических показателей, основанных на цитируемости* // Управление большими системами. – 2013. – № 44. – С. 248–261.
8. GUNS R., ROUSSEAU R. *Real and rational variants of the h-index and the g-index* // Journal of Informetrics. – 2009. – Vol. 3, №11. – P. 64–71.
9. HIRSCH J.E. *An index to quantify an individual's scientific research output* // Proc. National Academy of Sciences of the USA. – 2005. – Vol. 102, №46. – P. 16569–16572.
10. HIRSCH J.E. *An index to quantify an individual's scientific research output that takes into account the effect of multiple co-authorship* // Scientometrics. – 2010. – Vol. 85. – P. 741.

**ALTERNATIVE FRACTIONAL MODIFICATION OF
HIRSCH INDEX TO CARE FOR AUTHORS' COUNT IN
ARTICLE CITED**

Sergey Marvin, Ural Federal University named after the first President of Russia B.N. Yeltsin, Ekaterinburg, Candidate of Science, associate professor (s.v.marvin@yandex.ru).

Abstract: Shortcomings are studied of the Hirsch index and of the other similar scientometric indices, which do not consider co-authors' count or consider it with logical errors. The fractional analog is proposed of the Hirsch index, which cares for the co-authors' count and accurately measures the individual scientific merit of researchers from fields with traditionally low citation rates.

Keywords: scientometric, citation, Hirsch index, co-authorship, Hirsch core.

*Статья представлена к публикации
членом редакционной коллегии Д.А. Новиковым*

*Поступила в редакцию 09.04.2015.
Опубликована 31.07.2015.*