

СЕРГЕЙ ВЛАДИМИРОВИЧ МАРВИН

кандидат физико-математических наук, доцент кафедры
вычислительных методов и уравнений математической физики
Института радиоэлектроники и информационных технологий — РГФ
Уральского федерального университета
имени первого Президента России Б. Н. Ельцина,
Екатеринбург, Россия;
e-mail: s.v.marvin@yandex.ru



Нормированная доленая цитируемость как универсальная характеристика научной публикации

Предложена новая величина, характеризующая среднюю цитируемость научных работ по конкретной или собирательной тематике, относящихся к одному виду научных публикаций и одному году опубликования; в отличие от обычного среднего числа цитирований, предложенная величина учитывает количество авторов в публикациях, и поэтому названа средней доленой цитируемостью. Введено понятие нормированной доленой цитируемости, которая является отношением доленой цитируемости публикации к средней доленой цитируемости. Также предложен дробный аналог индекса Хирша, при вычислении которого следует использовать нормированную доленую цитируемость; вычисление индекса продемонстрировано на конкретных примерах.

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

Введение

В качестве основного наукометрического показателя отдельно взятой научной публикации используется количество ее цитирований. Однако этот показатель, считающийся в наукометрии базовым, не является исчерпывающим. Неоднократно отмечалось, что количество цитирований конкретной публикации совершенно неинформативно, если не указано время, прошедшее с момента ее появления, или не указана область исследований, к которой публикация относится (Маршакова-Шайкевич, 2013; Цыганов, 2013; Писляков, 2011). Также высказывалась мысль, что если речь идет о журнальной статье, то для правильной оценки количества ее цитирований необходимо уточнить, является ли данная статья оригинальной или обзорной (Писляков, 2011).

Бесспорно, обзорная статья не должна считаться более успешной публикацией в смысле цитируемости только потому, что на нее, как правило, приходится ссылок больше, чем на оригинальную статью. Обзорная статья — отдельный вид научной работы, и ее цитируемость будет корректно сравнивать только с цитируемостью других обзорных статей. Представляется необходимым распространить этот вывод на все другие разновидности научных публикаций для наукометрических баз, которые, в отличие от Web of Science, содержат сведения не только о журнальных статьях: недопустимо сравнивать цитируемость научных работ разного вида. Например, некорректно сравнивать цитируемость материалов конференций с цитируемостью статей, равно как сравнивать цитируемость материалов конференций или статей с цитируемостью научных монографий.

Итак, оценивать и сравнивать по числу цитирований корректно только научные работы с приблизительно одинаковым временем выхода в свет, относящиеся к одной области исследований и одному виду научных публикаций. Для таких научных работ имеет смысл вычислять среднее количество цитирований (SCImago Journal & Country Rank) и использовать его в качестве наукометрического «эталона успешности» публикации: определять для каждой научной работы, ниже ли ее цитируемость среднего значения или нет. Можно вычислять отношение числа цитирований конкретной статьи к средней цитируемости: если этот показатель меньше 1, то цитируемость ниже средней; если не меньше 1 — не ниже средней (в частности, если строго больше 1, то цитируемость выше средней). Деление конкретных значений величин на их эталонное значение в математике и физике принято называть нормировкой, поэтому указанный показатель цитируемости статьи уместно назвать нормированной цитируемостью.

Также существует подход к оценке публикационной активности и цитируемости ученого, основанный на делении его индекса Хирша h (Hirsch, 2005) за длительный отрезок времени на среднее число цитирований статей по тематике ученого, взятое по тому же отрезку времени (Цыганов, 2013). Однако такой подход объединяет при усреднении статьи с существенно различающимся временем опубликования, и непонятно, как его применять, когда ученый имеет публикации по разным научным направлениям (а такие ситуации встречаются).

Один из важных наукометрических выводов заключается в том, что наряду со временем, прошедшим с момента выхода в свет научной публикации, с областью исследований и видом публикации, необходимо также учитывать соавторство в научной работе, в частности количество авторов публикации. Самый простой способ учета соавторства — деление числа цитирований на количество авторов (Akhabue, Lautenbach, 2010; Михайлов, 2014; Полянин, 2014). Так возникает понятие долевого числа цитирований и долевого цитирования (Михайлов, 2014).

В литературе также предлагаются более сложные варианты распределения цитирований между соавторами (Hagen, 2013; Prathap, 2011; Михайлов, 2015). Но к устоявшейся практике формулы неравномерного распределения цитирований, по всей видимости, неприменимы: по сложившимся традициям многих научных журналов авторы статей располагаются в заголовках не в порядке убывания их значимости для публикации, а по алфавиту (что не является инициативой авторов). Изменение этих традиций в обозримом будущем маловероятно. И даже если это изменение произойдет, в сформированные наукометрические базы внести изменения будет физически невозможно из-за существенного объема данных, которые нужно будет исправить. Кроме того, в некоторых научных коллективах принято в качестве первого автора указывать первого по значимости, а всех остальных — по алфавиту без подробного анализа научного вклада каждого из них (Иванов, 2014). Идея отследить эти нюансы по сформировавшимся наукометрическим базам, к сожалению, нереализуема. Поэтому любые предложения по неравномерному распределению цитирований между авторами, в лучшем случае, относятся к далекому будущему. Причем даже если эти предложения будут реализованы в рамках какой-либо наукометрической базы, они будут применимы к научным работам, вышедшим в свет только после переломного момента, когда все журналы без исключения позволят перечислять фамилии авторов в заголовке статей на их авторское усмотрение, и авторы будут предупреждены о новых правилах наукометрической базы. Но до тех пор в качестве универсальной формулы следует использовать простое деление числа цитирований на число авторов.

Явление соавторства находит отражение в некоторых модификациях индекса Хирша (Batista, Campiteli, Kinouchi, Martinez, 2006; Hirsch, 2010; Publish or Perish; Schreiber, 2008; Михайлов, 2014; Цыганов, 2013; Штовба, 2013). С использованием долевой цитируемости вычисляется аналог h , который учитывает соавторство и при этом наиболее близок к индексу Хирша по своему определению (Publish or Perish; Михайлов, 2014; Цыганов, 2013). Он называется Individual h -index (PoP variation) (Publish or Perish; Цыганов, 2013); также для него было предложено название «модифицированный индекс Хирша» и краткое обозначение h^* (Михайлов, 2014). Правило, по которому вычисляется h^* , можно сформулировать следующим образом: h^* ученого равен k , если у него есть k научных работ, долевое число цитирований каждой из которых не меньше k и при этом у каждой из остальных его работ долевое число цитирований не превосходит k . Рассмотрены конкретные примеры вычисления h^* ; показано, как переход от обычного индекса Хирша к индексу h^* может изменить картину публикационной активности и цитируемости ученого (Михайлов, 2014).

Индекс h^* , подобно обычному индексу Хирша, является целочисленным показателем, скачкообразно изменяющимся от одного целого значения к другому. Как следствие, у него очень медленная динамика в областях науки с традиционно невысокой цитируемостью или большими исследовательскими коллективами. Поэтому была предложена альтернативная дробная модификация h^* , более чувствительная к малым дробным изменениям долевой цитируемости; соответствующий модифицированный индекс Хирша был обозначен как h_{mod} (Марвин, 2015).

Введение в рассмотрение долевой цитируемости требует определенной модификации понятий средней и нормированной цитируемости. В данной статье будет установлено, какую величину следует взять в качестве средней долевой цитируемости. Также будет предложена дробная модификация индекса Хирша, в вычислении которой задействована нормированная, а не обычная долевая цитируемость. Будут разобраны конкретные примеры.

Определение средней долевой цитируемости

В математике для различных целей используются различные варианты усреднения величин. Для корректного определения средней долевой цитируемости необходимо тщательно разобрать смысл неусредненной долевой цитируемости. Предположим, что научная работа имеет C цитирований и n авторов; ее долевая цитируемость $c = C/n$. Это означает, что при наукометрическом анализе данная публикация отождествляется с набором из n публикаций, написанных авторами единолично; каждая из этих n публикаций имеет C/n цитирований, то есть по c цитирований каждому автору. Для научных работ, написанных единолично, среднее число цитирований, очевидно, должно вычисляться обычным способом: сумма цитирований всех работ делится на общее количество этих работ. Если же мы рассматриваем M работ, написанных единолично или в соавторстве, причем i -я публикация процитирована C_i раз и написана n_i авторами (номер i меняется в нашей выборке публикаций от 1 до M), то общее количество цитирований единолично написанных работ, на которые условно разбиваются публикации из рассматриваемой выборки, определяется следующим образом:

$$c_1 \cdot n_1 + c_2 \cdot n_2 + \dots + c_M \cdot n_M = C_1 + C_2 + \dots + C_M,$$

то есть, это общее количество цитирований, вычисляемое обычным способом.

Количество единолично написанных работ, на которые условно разделены публикации:

$$n_1 + n_2 + \dots + n_M,$$

то есть равно суммарному количеству авторов всех рассматриваемых публикаций. Следовательно, универсальная формула для вычисления средней долевой цитируемости, которую мы обозначим как \bar{c} , имеет следующий вид:

$$\bar{c} = \frac{C_1 + C_2 + \dots + C_M}{n_1 + n_2 + \dots + n_M}, \quad (1)$$

то есть общее количество цитирований рассматриваемых научных работ следует разделить на общее количество их авторов. При поверхностном рассмотрении понятия долевой цитируемости можно было предположить, что в качестве средней долевой цитируемости следует взять сумму долевого числа цитирования отдельных работ c_1, c_2, \dots, c_M и разделить ее на общее количество работ M . Такой подход был бы неверным.

Выполним тождественные преобразования формулы (1):

$$\bar{c} = \frac{C_1 + C_2 + \dots + C_M}{n_1 + n_2 + \dots + n_M} = \frac{(C_1 + C_2 + \dots + C_M) \cdot M}{(n_1 + n_2 + \dots + n_M) \cdot M} = \frac{\bar{C}}{\bar{n}}, \quad (2)$$

где $\bar{C} = (C_1 + C_2 + \dots + C_M)/M$ — среднее число цитирований (не долевое); $\bar{n} = (n_1 + n_2 + \dots + n_M)/M$ — среднее число авторов. Выражение (2) раскрывает еще один смысл формулы (1): среднее долевое число цитирований равно отношению среднего числа цитирований к среднему количеству авторов.

Нормированную долевую цитируемость i -й работы обозначим $c_{\text{norm}, i}$ и определим следующим образом: $c_{\text{norm}, i} = 0$, если $c_i = 0$; если же $c_i \neq 0$, то $c_{\text{norm}, i}$ вычисляется по формуле:

$$c_{\text{norm}, i} = \frac{c_i}{\bar{c}}. \quad (3)$$

Отдельное пояснение, касающееся $c_i = 0$, было необходимо, потому что в конкретной выборке публикаций не только i -я, но и все другие публикации могут иметь нулевое число цитирований; это приведет к нулю в знаменателе в формуле (3), то есть формула (3) будет неприменимой. Поэтому отдельно оговорено, что $c_{\text{norm}, i} = 0$ при $c_i = 0$, даже если i -я публикация, имеющая нулевое число цитирований, такая же по цитируемости, как все остальные работы в рассматриваемой выборке объема M .

Точность определение времени опубликования научной работы, ее области исследований и вида для вычисления средних показателей в рамках Elibrary.ru

Во введении отмечено, что корректно сравнивать и вычислять средние значения только для публикаций, вышедших в свет приблизительно в одно и то же время. Для конкретных расчетов необходимо определить, как следует понимать эту приблизительность. В РИНЦ (Российском индексе научного цитирования), то есть в электронной научной базе Elibrary.ru функция расширенного поиска позволяет находить публикации по году опубликования, но не по меньшему периоду времени. Причем функция упорядочения по дате выпуска работает исключительно по годам: публикации одного и того же года при использовании этой функции в хронологическом порядке не располагаются.

Если при сравнении и усреднении цитируемости рассматриваются научные работы одного года опубликования, то необходимо сгладить различие между началом и окончанием года: следует определять цитируемость публикаций и проводить вычисления по формуле (1) не сразу после окончания года, а по прошествии достаточно длительного промежутка времени. В данной статье в качестве компромисса между достаточной длительностью периода ожидания и оперативностью наукометрических исследований предлагается трехлетний период. То есть принимать во внимание цитируемость, например, публикаций 2012 года (для их сравнения и усреднения) следует начиная только с 2016 года, чтобы различие между январем и декабрем 2012 года было небольшим в относительных величинах. Для оперативности наукометрических расчетов можно взять полугодия и полуторагодовой период ожидания (однако при самостоятельных исследованиях это сделать практически невозможно из-за специфики программного обеспечения РИНЦ). Уменьшать указанные периоды и добиваться еще большей оперативности не следует, так как это может привести к недопустимо малой выборке научных публикаций. Кроме того, большей оперативности достичь будет затруднительно из-за длительности обработки запросов на добавление публикаций в Elibrary.ru. Далее в рамках данной статьи будем считать, что научные работы следует объединять по годам опубликования.

Для корректности сравнения и усреднения необходимо, чтобы публикации рассматриваемой выборки относились к одной тематике. В РИНЦ тематика научных работ определяется кодами ГРНТИ (Государственным рубрикатором научнотехнической информации), причем, как правило, у любой публикации не один, а большее число таких кодов. Возникает нетривиальная проблема, как для данной конкретной публикации определить выборку научных работ, которым она соответствует по тематике, чтобы по этой выборке вычислить среднюю долевую цитируемость, а для рассматриваемой публикации — нормированное доленое число цитирований. Если публикация относится к тематикам A_1, A_2, \dots, A_L , то, на первый взгляд, следует взять их пересечение: включить в выборку для усреднения те публикации, которые тоже относятся ко всем этим тематикам одновременно, и никакие другие работы в выборку не включать. Функция расширенного поиска Elibrary.ru позволяет это сделать: сначала формируется поисковый запрос работ тематики A_1 , затем с использованием функции «Искать в результатах предыдущего запроса» формируется новый поисковый запрос на тематику A_2 , и так далее до A_L . Однако применение этого приема к конкретным наборам тематик показывает, что объем получающейся

выборки научных работ оказывается недопустимо малым. Как следствие, возникает вопрос об осмысленности вычисляемой средней доленой цитируемости. Нередки ситуации, когда публикация с данным набором тематик оказывается уникальной в своем роде.

Вообще, для статистических исследований представляется недопустимым дробление мелких тематик, описываемых кодами ГРНТИ третьего уровня (шести-значными кодами ГРНТИ), на еще более мелкие подразделы. Поэтому для формирования выборки публикаций предполагается целесообразным использовать не пересечение, а объединение тематик: объединить в одну выборку публикации, относящиеся к A_1 ; публикации, относящиеся к A_2 , и так далее до A_L . При формировании поискового запроса объединение тематик задается простым перечислением кодов ГРНТИ в поле «Тематика».

Для корректности сравнения и усреднения также необходимо, чтобы работы относились к одному виду научных публикаций. В функции расширенного поиска РИНЦ научные публикации разделены на следующие виды: статьи в журналах, книги, материалы конференций, депонированные рукописи, диссертации, отчеты и патенты. Следует провести и дальнейшее разделение.

Оригинальные статьи, обзорные статьи и рецензии необходимо рассматривать в отдельных выборках. Однако краткие сообщения, которые неслучайно называются в некоторых журналах краткими статьями, будет корректно рассматривать вместе с оригинальными статьями: это публикации того же вида, только небольшие по объему. В Elibrary.ru также включены редакторские заметки информационного и поздравительного характера. Размещение таких материалов в журналах не вызывает возражений, однако в базе научного цитирования сведения о таких статьях неуместны, и при наукометрических исследованиях эти заметки должны исключаться из рассмотрения.

Из книг, размещенных в РИНЦ, по всей видимости, должны быть исключены учебники и учебные пособия: необходимость учебно-методической работы в высших учебных заведениях очевидна, однако учебно-методическая работа и научно-исследовательская работа — разные стороны публикационной деятельности преподавателя.

Если исключить учебники и учебно-методические пособия, то оставшиеся книги следует разделить на монографии и справочники.

В разделе «Диссертации» при сравнении и усреднении цитируемости необходимо разделять сами диссертации и их авторефераты. Заметим, что для публикаций этого вида доленая цитируемость совпадает с обычной, так как диссертации и авторефераты пишутся единолично.

В подробных описаниях публикаций Elibrary.ru вид публикаций указан максимально конкретно: оригинальная ли данная статья или обзорная, является ли данная книга справочником или монографией и т. д. Однако указанные разграничения не проведены в функциях расширенного поиска, и это затрудняет не только внешние наукометрические исследования, но и вообще любой поиск в РИНЦ.

Теперь можно окончательно сформулировать правила, по которым следует определять для конкретной публикации ее нормированную доленую цитируемость. Предположим, научная работа опубликована в N -м году и относится к тематикам A_1, A_2, \dots, A_L (анализ ее цитируемости проводится $(N + 4)$ -м году или позже). Тогда мы объединяем в одну выборку все публикации тематик A_1, A_2, \dots, A_L , относящиеся-

ся к тому же виду научных работ (оригинальные статьи, или обзорные статьи, или монографии, и т. д.) и вышедшие в свет в N -м году. Для этой выборки находим среднюю долевую цитируемость по формуле (1); затем для рассматриваемой публикации находим нормированную долевую цитируемость по формуле (3).

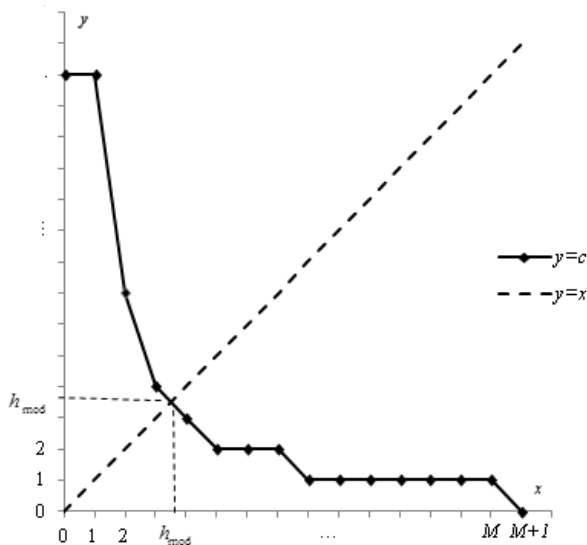
Например, если оригинальная статья 1996 года относится к тематике «Испытание материалов. Дефектоскопия» (код ГРНТИ: 81.09.81) и к тематике «Магнитные материалы» (код ГРНТИ: 45.09.29), то для вычисления ее нормированной долевой цитируемости мы собираем в одну выборку все оригинальные статьи 1996 года по дефектоскопии и по магнитным материалам. Для получившейся выборки вычисляем среднюю долевую цитируемость, затем находим нормированную долевую цитируемость рассматриваемой статьи. Прибегая к аналогиям из техники, можно сказать, что данная статья «открывает шлюз» между тематиками «Дефектоскопия» и «Магнитные материалы» (в рамках 1996 года), перемешивает статьи по этим тематикам и устанавливает некоторое среднее значение долевой цитируемости объединенной тематики.

Заметим, что нормированная долевая цитируемость может служить универсальным показателем цитируемости научной публикации: по ней допустимо сравнивать научные работы, относящиеся к разным областям исследования или к разным видам публикаций.

Дробная модификация индекса Хирша, вычисляемая через нормированную долевую цитируемость

Дробный индекс h_{mod}^* , призванный скомпенсировать медленную динамику h^* в областях науки с традиционно невысокой цитируемостью, имеет графическое определение (Марвин, 2015). Предположим ученый, публикационную активность и цитируемость которого мы оцениваем, имеет M научных работ. Следует пронумеровать эти работы в порядке невозрастания их долевой цитируемости: $c_1 \geq c_2 \geq \dots \geq c_M$. Затем нужно отметить на графике точки с координатами $(1; c_1)$, $(2; c_2)$, ..., $(M; c_M)$ и соединить их ломаной (рис. 1). Эта ломаная дополняется слева еще одной вершиной: $(0; c_0) = (0; c_1)$. Таким образом, первый отрезок ломаной в любом случае будет горизонтальным. Можно сказать, что вводится «мнимая» нулевая публикация, долевое число цитирований которой равно долевого числу цитирований первой публикации. Если долевое число цитирований последней M -й публикации больше 0, то вводится «мнимая» $(M+1)$ -я публикация, долевое число цитирований которой равно 0. Таким образом, ломаная дополняется наклонным звеном с вершинами $(M; c_M)$ и $(M+1; 0)$. Заметим, что при анализе цитируемости ученого наличие публикации с нулевой цитируемостью равносильно ее отсутствию.

Для определения индексов h^* и h_{mod}^* необходимо найти точку пересечения ломаной и прямой $y = x$ (рис. 1). Индекс h^* равен целой части абсциссы точки пересечения (равно как и ординаты: точка располагается на прямой $y = x$). Индекс h_{mod}^* равен этой абсциссе без взятия целой части. h_{mod}^* , который обычно оказывается дробным, очевидно, более чувствителен к перемещению звена ломаной, с которым пересекается прямая, в частности к изменению его наклона. При введении «мнимых» публикаций ломаная соединяется с осями координат. Тем самым гарантируется

Рис. 1. Графическое определение h_{mod}

пересечение прямой и ломаной, даже если первая вершина ломаной располагается слишком низко (под прямой $y = x$) или M -я вершина располагается слишком высоко (над прямой $y = x$).

Индекс h_{mod} подробно исследован, его замысел полностью аргументирован, разобраны конкретные примеры и получена точная формула для вычисления (Марвин, 2015):

$$h_{mod} = \frac{c_n \cdot (n+1) - c_{n+1} \cdot n}{c_n + 1 - c_{n+1}}, \quad (4)$$

где номера публикаций n и $n+1$ определяются так, что долевое число цитирований n -й публикации $c_n \geq n$ и при этом долевое число цитирований $(n+1)$ -й публикации $c_{n+1} \leq n+1$. На графике n и $n+1$ — абсциссы вершин ломаной, между которыми заключена точка пересечения ломаной с прямой $y = x$.

Если вместо долевого цитируемости использовать нормированную долевую цитируемость, идея дробного аналога индекса Хирша оказывается еще более востребованной: следует ожидать, что у большинства научных работ нормированная долевая цитируемость не превосходит 2. Это означает, что целочисленный аналог индекса Хирша, при вычислении которого будет использоваться нормированная долевая цитируемость, у большинства ученых примет значения 0, 1 и 2. Следовательно, такой индекс непригоден для ранжирования ученых по их публикационной активности и цитируемости. Дробную модификацию индекса Хирша, связанную с нормированной долевой цитируемостью, обозначим как h_{norm} и будем вычислять по формуле, аналогичной (4):

$$h_{norm} = \frac{c_{norm,n} \cdot (n+1) - c_{norm,n+1} \cdot n}{c_{norm,n} + 1 - c_{norm,n+1}}, \quad (5)$$

где $c_{norm,n} \geq n$ и $c_{norm,n+1} \leq n+1$.

Так как при вычислении $h_{\text{норм}}$ используется нормированная долевая цитируемость, этот индекс можно считать универсальным и по его значению сравнивать публикационную активность и цитируемость ученых, задействованных в разных областях науки. Однако практическая значимость этого индекса заключается не только в этом.

Как отмечалось во введении, научные интересы некоторых авторов невозможно отнести к какой-либо конкретной тематике; причем невозможно даже если тематика будет определяться кодами ГРНТИ не третьего, а второго уровня (второму уровню соответствуют четырехзначные коды ГРНТИ, указывающие область научных исследований не так конкретно, как шестизначные коды ГРНТИ). Двухзначные коды ГРНТИ первого уровня характеризуют научные публикации слишком обобщенно (например, просто «Математика» с кодом ГРНТИ 27 или просто «Физика» с кодом ГРНТИ 29); но даже при такой общности встречаются ученые, совокупность научных работ которых не соответствует никакому конкретному коду ГРНТИ первого уровня.

Ученые, занятые в междисциплинарных исследованиях, нередко являются авторами статей, относящихся к различным наборам тематик. Например, исследователи, занимающиеся применением математических методов в дефектоскопии, являются в основном авторами публикаций, находящихся в пересечении дефектоскопии и математической физики, но иногда пишут статьи, посвященные исключительно математическим проблемам. Руководители научных коллективов, специализирующихся на квантовой теории твердых тел, занимаются исследованиями как по сверхпроводимости, так и по магнетизму (и поэтому становятся соавторами статей, посвященных различным научным проблемам). Кроме того, нередки ситуации, когда ученый радикально меняет область своих научных интересов (например, после защиты диссертации). Заметим также, что в различных областях науки встречаются ученые, которые интересуются наукометрической тематикой и пишут по ней статьи.

Вычисление h ученого предполагает использование массива всех его публикаций, сведения о которых имеются в наукометрической базе. Однако недопустимо рассматривать в одном ряду статьи, монографии, тезисы и научные публикации всех других видов, одинаковым образом их учитывая при анализе публикационной активности и цитируемости. С другой стороны, при таком анализе крайне нежелательно оставлять публикации только одного вида (например, журнальные статьи), полностью игнорируя работу ученого по другим публикациям.

В вычислении $h_{\text{норм}}$ задействована нормированная долевая цитируемость, и таким образом специфика разных областей исследования и разных видов научных работ оказывается автоматически учтенной. Это делает допустимым при вычислении индекса совокупное использование всех без исключения публикаций ученого, даже если это публикации разного вида или относятся к разным областям исследований.

В наукометрических базах все дробные показатели вычисляются с точностью до какого-то определенного количества знаков за запятой. Поэтому необходимо определиться с точностью вычисления $h_{\text{норм}}$. Как отмечалось выше, следует быть готовыми к тому, что нормированная долевая цитируемость большинства научных работ будет находиться в числовом отрезке между 0 и 2. Как следствие, в этих же пределах будет колебаться $h_{\text{норм}}$ большинства ученых. Ввиду этого обстоятельства представляется целесообразным вычислять $h_{\text{норм}}$ с точностью до двух знаков за запятой:

в отрезке от 0 до 2 будет двести одно значение этого индекса, что вполне достаточно для ранжирования ученых по их публикационной активности и цитируемости (без каких-либо чрезмерных уравниваний авторов научных публикаций).

$c_{\text{norm}, i}$, как промежуточный результат вычислений, будем определять с большей точностью — до трех знаков. Тогда показатели c_i и \bar{c} , используемые при вычислении $c_{\text{norm}, i}$, следует вычислять с точностью до четырех знаков. Заметим, что такая точность вычисления долевого цитируемости достаточна для того, чтобы увеличение количества цитирований на 1 было заметным даже у статей, имеющих рекордное количество авторов (Клейн, 2015).

Примеры вычисления h_{norm}

Вычислим предложенный индекс h_{norm} двух реальных ученых. Их фамилии не раскрываются; обозначать их будем как S_1 и S_2 . Все необходимые вычисления проводились в программе Excel. При исследовании публикационной активности и цитируемости выдержан трехлетний период ожидания: исследования проводились в августе 2015 года, и при этом рассматривались научные работы не позже 2011 года (хотя заметим, что более поздние работы S_1 и S_2 все равно пока не цитировались, а на величину h_{norm} такие работы не влияют).

Ученый S_1 работает в области неразрушающего контроля. Все его публикации относятся к тематике «Испытание материалов. Дефектоскопия», причем большая часть их относится также и к другим тематикам: к «Теории магнитных свойств твердых тел» (код ГРНТИ: 29.19.37), или к «Приборам неразрушающего контроля изделий и материалов» (код ГРНТИ: 59.45), или к «Магнитным материалам». Работы S_1 , число цитирований которых отлично от 0, — это десять оригинальных статей и автореферат докторской диссертации.

Следует отметить некоторые трудности, связанные с техническими ограничениями базы РИНЦ. Во-первых, в описаниях публикаций коды ГРНТИ не указаны. Эти коды можно определить только методом проб и ошибок, задавая в расширенном поиске Elibrary.ru в поле «Авторы» фамилию и инициалы ученого, а в поле «Тематика» — предполагаемый код ГРНТИ работы. При таком подходе некоторые тематики, к которым отнесена публикация, могут быть потеряны. Но это касается только тех тематик, к которым работа может быть отнесена совершенно неожиданно; так что, скорее всего, при таком вынужденном способе определения кода ГРНТИ не вносится погрешность, а исключаются недоразумения.

Также следует заметить, что в результатах расширенного поиска по году и тематике оригинальные и переводные версии русскоязычных статей не объединяются, а идут как отдельные публикации (они объединяются только в авторских профилях, но не в результатах расширенного поиска). Это очень сильно затрудняет исследование. Кроме того, затрудняет поиск оригинальных статей по данной тематике недостаток базы, отмеченный выше: оригинальные статьи идут в результатах поиска совместно с обзорными статьями, рецензиями и редакторскими заметками. Обзорные статьи определяются отдельным поиском: в поле опознавательных слов «Что искать» задается слово *обзор* или *review*, далее по подробным описаниям найденных публикаций определяется, действительно ли они являются обзорными статьями.

Число их цитирований и суммарное количество их авторов вычитаются из числителя и знаменателя формулы (3). Рецензии и редакторские заметки легко распознаются по названию в общих результатах поиска.

Количество авторов научной работы в результатах поиска не указывается отдельным столбцом, поэтому авторов приходится считать вручную, вводя полученные числа в программу Excel. Также в качестве недостатка следует отметить, что в результатах поиска указаны все цитирования статей, в том числе и цитирования авторами. Однако было бы интересно рассчитать h_{norm} без учета самоцитирований и цитирований соавторами. Для конкретного автора ручной отбор самоцитирований и соавторских цитирований не представляет больших трудностей. Но тогда среднюю долевую цитируемость по формуле (1) следует считать тоже без учета цитирований статей их авторами. Выборки статей по годам настолько большие, что ручной отбор цитирований представляется физически невозможным. Поэтому h_{norm} в рамках проведенных исследований рассчитан с учетом самоцитирований и цитирований соавторами.

В таблице 1 приведены данные по тем публикациям ученого S_j , которые имеют ненулевое число цитирований. Каждый столбец таблицы, начиная со второго, соответствует конкретной публикации автора. Знаком ΣC_i обозначена сумма цитирований всех публикаций, использующихся для вычисления \bar{c} (выборка публикаций составляется по сформулированным выше принципам). Знаком Σn_i обозначена соответствующая сумма авторов.

Таблица 1

C	3	1	4	3	3	2	1	3	1	1	1
n	4	4	2	2	1	3	4	3	2	2	5
c	0,7500	0,2500	2,0000	1,5000	3,0000	0,6667	0,2500	1,0000	0,5000	0,5000	0,2000
ΣC_i	149	859	8807	8807	31	643	721	2098	15590	2631	2242
Σn_i	245	1383	6555	6555	14	543	767	3046	18399	3676	3860
\bar{c}	0,6082	0,6211	1,3436	1,3436	2,2143	1,1842	0,9400	0,6888	0,8473	0,7157	0,5808
c_{norm}	1,233	0,403	1,489	1,116	1,355	0,563	0,266	1,452	0,590	0,699	0,344

Долевые цитируемости работ S_j , перечисленные в порядке невозрастания: 3; 2; 1,5; 1; 0,75; 0,6667; 0,5; 0,5; 0,25; 0,25; 0,2. На втором месте этой выборки располагается число 2, и в такой ситуации нет необходимости применять формулу (4): $h_{\text{mod}} = 2$. Нормированные долевые цитируемости, перечисленные в порядке невозрастания: 1,489; 1,452; 1,355; 1,233; 1,116; 0,699; 0,59; 0,563; 0,403; 0,344; 0,266. На первом месте этой выборки располагается 1,489, что превосходит 1, а на втором месте — 1,452, что меньше 2. Следовательно, h_{norm} определяется по первым двум публикациям (заметим, что первые пары публикаций в двух разных выборках — разные, так как публикации в выборках упорядочены по разным величинам). Применяв формулу (5), получаем $h_{\text{norm}} = 1,47$.

Ученый S_2 — математик; конкретные тематики: «Дифференциальные уравнения с частными производными» (код ГРНТИ: 27.21); «Обобщенные функции» (код ГРНТИ: 27.39.17); «Теория вероятностей и случайные процессы» (код ГРНТИ: 27.43.15). В работах S_2 перечисленные тематики встречаются как по отдельности,

так и в различных сочетаниях. В таблице 2 приведены данные по публикациям с нулевой цитируемостью этого ученого; в числе таких публикаций — три оригинальные статьи и отчет по научно-исследовательской работе.

Таблица 2

C	1	1	1	4
n	6	1	1	2
c	0,1667	1,0000	1,0000	2,0000
ΣC_i	6	2389	598	512
Σn_i	35	1406	350	744
\bar{c}	0,1714	1,6991	1,7086	0,6882
c_{norm}	0,972	0,589	0,585	2,906

Долевые цитируемости статей S_2 , перечисленные в порядке невозрастания: 2; 1; 1; 0,1667. На первом месте этой выборки располагается 2 (число, большее 1), а на втором — 1 (число, меньшее 2). После применения формулы (4) получаем $h_{\text{mod}} = 1,5$. Нормированные долевые цитируемости, перечисленные в порядке невозрастания: 2,906; 0,972; 0,589; 0,585. На первом месте этой выборки располагается число 2,906, которое больше 1, а на втором месте располагается 0,972, что меньше 2. Следовательно, h_{norm} определяется по первым двум публикациям, и после применения формулы (5) получаем $h_{\text{norm}} = 1,65$. То есть величина h_{mod} больше у S_1 , но h_{norm} , учитывающий специфику научных публикаций, выше у S_2 .

Заключение

Предложенная в данной статье величина нормированной долевой цитируемости является универсальной характеристикой, безотносительной к области исследований и виду публикации; в этом заключается ее практическая значимость. Вычисление нормированной долевой цитируемости длительно и трудоемко, если используется только программа Excel. Однако программные возможности любой наукометрической базы достаточны для того, чтобы вычисление этой величины было полностью автоматизировано и исчислялось секундами: ориентировочно такое время требуется наукометрической базе мирового уровня для вычисления средних показателей недолевой цитируемости и построения графиков распределения по годам и странам в (SCImago Journal & Country Rank).

Литература

- Akhabue E., Lautenbach E.* Equal contributions and credit: An emerging trend in the characterization of authorship // *Annals of Epidemiology*. 2010. Vol. 20. № 11. P. 868–871.
- Baïsta P. D., Campiteli M. G., Kinouchi O., Martinez A. S.* Is it possible to compare researchers with different scientific interests? // *Scientometrics*. 2006. Vol 68. № 1. P. 179–189.

Hagen N. T. Harmonic coauthor credit: A parsimonious quantification of the byline hierarchy // Journal of Informetrics. 2013. Vol. 7. № 4. P. 784–791.

Hirsch J. E. An index to quantify an individual's scientific research output // Proceeding of National Academy of Sciences of the USA. 2005. Vol. 102. № 46. P. 16569–16572.

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. № 2. P. 741–754.

Publish or Perish. URL: <http://www.harzing.com/pop.htm#metrics>.

Prathap G. The fractional and harmonic p-indices for multiple authorship // Scientometrics. 2011. Vol. 86. № 2. P. 239–244.

Schreiber M. A modification of the h-index: The h(m)-index accounts for multi-authored manuscripts // Journal of Informetrics. 2008. Vol. 2. № 3. P. 211–216.

SCImago Journal & Country Rank. URL: <http://www.scimagojr.com/index.php>

Иванов К. С. Индекс Хирша: модернизация необходима (еще раз о количественной оценке работы ученых) // Троицкий вариант — Наука. Рубрика: Гайд-парк онлайн. 30.07.2014. URL: <http://trv-science.ru/2014/07/30/indeks-khirsha-modernizaciya-neobkhodima> [*Ivanov K. S.* Indeks Khirsha: modernizatsiya neobkhodima (yeshche raz o kolichestvennoy otsenke raboty uchenykh) // Troitskiy variant — Nauka. Rubrika: Gayd-park onlayn. 30.07.2014. URL: <http://trv-science.ru/2014/07/30/indeks-khirsha-modernizaciya-neobkhodima>].

Клейн Л. С. Драма соавторства // Троицкий вариант — Наука. Рубрика: Авторские колонки. 30.07.2015. № 182. [*Kleyn L. S.* Drama soavtorstva // Troitskiy variant — Nauka. Rubrika: Avtorskiye kolonki. 30.07.2015. № 182.]

Марвин С. В. Альтернативная дробная модификация индекса Хирша, учитывающая количество авторов цитируемых статей // Управление большими системами. 2015. Вып. 56. С. 108–122. [*Marvin S. V.* Al'ternativnaya drobnaya modifikatsiya indeksa Khirsha, uchityvayushchaya kolichestvo avtorov tsitiruyemykh statey // Upravleniye bol'shimi sistemami. 2015. Vyp. 56. S. 108–122.]

Маршакова-Шайкевич И. В. Роль библиометрии в оценке исследовательской активности науки // Управление большими системами. 2013. Вып. 44. С. 210–247. [*Marshakova-Shaykevich I. V.* Rol' bibliometrii v otsenke issledovatel'skoy aktivnosti nauki // Upravleniye bol'shimi sistemami. 2013. Vyp. 44. S. 210–247.]

Михайлов О. В. О возможной модификации индексов Хирша и Эгга с учетом соавторства // Социология науки и технологий. 2014. Т. 5. № 3. С. 48–56. [*Mikhaylov O. V.* O vozmozhnoy modifikatsii indeksov Khirsha i Egga s uchetom soavtorstva // Sotsiologiya nauki i tekhnologiy. 2014. T. 5. № 3. S. 48–56.]

Михайлов О. В. Новая версия h-индекса с учетом числа соавторов и порядка их перечисления в научной публикации // Социология науки и технологий. 2015. Т. 6. № 2. С. 24–32. [*Mikhaylov O. V.* Novaya versiya h-indeksa s uchetom chisla soavtorov i poryadka ikh perechisleniya v nauchnoy publikatsii // Sotsiologiya nauki i tekhnologiy. 2015. T. 6. № 2. S. 24–32.]

Писляков В. В. Наука через призму статей // Публичные лекции «Полит.ру». 21.12.2011. URL: http://polit.ru/article/2011/12/21/pislyakov_2011 (дата обращения: 29.01.2016) [*Pislyakov V. V.* Nauka cherez prizmu statey // Publichnyye lektzii «Polit.ru». 21.12.2011. URL: http://polit.ru/article/2011/12/21/pislyakov_2011] (date accessed: 29.01.2016)

Полянин А. Д. Недостатки индексов цитируемости и Хирша и использование других наукометрических показателей // Математическое моделирование и численные методы. 2014. Вып. 1. С. 131–144. [*Polyanin A. D.* Nedostatki indeksov tsitiruyemosti i Khirsha i ispol'zovaniye drugikh naukometricheskikh pokazateley // Matematicheskoye modelirovaniye i chislennyye metody. 2014. Vyp. 1. S. 131–144.]

Цыганов А. В. Краткое описание наукометрических показателей, основанных на цитируемости // Управление большими системами. 2013. Вып. 44. С. 248–261. [*Tsyganov A. V.* Kratkoye opisaniye naukometricheskikh pokazateley, osnovannykh na tsitiruyemosti // Upravleniye bol'shimi sistemami. 2013. Vyp. 44. S. 248–261.]

Штовба С. Д., Штовба Е. В. Обзор наукометрических показателей для оценки публикационной деятельности ученого // Управление большими системами. 2013. Вып. 44. С. 262–278. [*Shtovba S. D., Shtovba Ye. V.* Obzor naukometricheskikh pokazateley dlya otsenki publikatsionnoy deyatel'nosti uchenogo // Upravleniye bol'shimi sistemami. 2013. Вып. 44. С. 262–278.]

Normalized shared citation as a universal characteristic of a scientific publication

SERGEY V. MARVIN

associate professor

Department of computational methods and equations of mathematical physics,

Institute of radioelectronics and information technologies — RTF,

Ural Federal University named after the first

President of Russia B. N. Yeltsin

Ekaterinburg, Russia;

e-mail: s.v.marvin@yandex.ru

The new value that characterizes an average citation of scientific papers on a particular subject or collective, relating to an identical type of publications research and equal year of publication, is proposed; unlike the usual average number of citations, take into account the value of the proposed number of authors in publications, so named average equity citation. A notion of normalized shared citation, which is the ratio of shared citation to the average shared citation, is defined. Also is proposed the fractional analog of the Hirsch index, the calculation of which must to use normalized shared citation; calculation of the index illustrated with specific examples.

Keywords: scientometrics, citation, co-authorship, Hirsch index.