

## Индекс цитирования, учитывающий скрытую диффузию научных знаний

*Предложен новый наукометрический показатель, который оценивает диффузию знаний и имеет две составляющие: первая эквивалентна обычному индексу цитирования, т. е. описывает видимую диффузию научных знаний; вторая – отражает скрытую диффузию научных знаний и выражается через количество неявных цитирований. Практическая ценность предложенного показателя заключается в том, что он позволяет легко идентифицировать скрытых инициаторов научного мейнстрима. Отличительная особенность таких ученых – большое значение предложенного индекса цитирования и малое значение обычного индекса цитирования.*

**Ключевые слова:** наукометрия, индекс цитирования, неявное цитирование, диффузия знаний, инициатор мейнстрима

### ВВЕДЕНИЕ

Для оценивания полезности научной деятельности сегодня совместно с экспертными заключениями все чаще используют и наукометрические показатели. В первую очередь наукометрические показатели востребованы для оценки фундаментальных исследований, результаты которых непосредственно не связаны с экономическим эффектом. Фундаментальные разработки направлены на развитие науки, поэтому их востребованность оценивают через отзыв научного сообщества на публикации с результатами исследований. Формально этот отзыв выражают индексом цитирования – суммарным количеством публикаций, ссылающихся на рассматриваемые работы. Варианты индекса цитирования модифицируют его следующим образом:

- учитывают личный вклад ученого [1, 2], разделяя количество цитирований между соавторами, число которых в одной статье иногда превышает 3000;
- учитывают авторитетность цитирующего издания, используя импакт-фактор журнала или другой аналогичный коэффициент [3, 4];
- игнорируют самоцитирование или цитирование соавторами [5, 6], что существенно снижает рейтинг "ученого-затворника", публикации которого интересуют только его самого;
- игнорируют повторное цитирование одной работы одним и тем же ученым [7], что уменьшает влияние комплементарного цитирования.

Использование этих и других модификаций не устраняют следующие два недостатка индекса цитирования. Первый недостаток связан с забыванием имен классиков, когда авторы считают, что вклад предшественников настолько хорошо известен любому из соответствующей области науки, что нет смысла об этом упоминать [8]. Соответственно, работы таких ученых высшего ранга быстро перестают

цитировать. Как правило, этот процесс ускоряет включение научных результатов классиков в учебники и учебные пособия. Поэтому стремительно возрастает диффузия этих научных знаний, т.е. их адаптация и использование в широком спектре научных и инженерных исследований и разработок [9].

Второй недостаток индекса цитирования заключается в низком рейтинге концептуальных работ с принципиально новыми идеями. Это происходит из-за достаточно распространенного сокрытия первоисточников, т.е. невключения концептуальных работ предшественников в перечень цитируемой литературы. Создание нового научного результата – это процесс, включающий этапы [8]: 1) генерирования идеи и разработки концепции, 2) модификации идеи (концепции) для улучшения точности, быстродействия, упрощения и т.д., 3) продвижения и практического использования. В истории науки есть много примеров, когда работы 2-го и 3-го этапов значительно чаще цитируют, чем концептуальные статьи. Впечатляющим примером, из приведенных в [8] являются статьи А. Фолина [10] и О. Лоури [11]. О. Лоури модифицировал реактив А. Фолина для колориметрического определения белка путем добавления еще одного компонента, что расширило границы применения метода. Сегодня, по данным Google Scholar, статью О. Лоури процитировали 247 480 раз, тогда как работу А. Фолина – 1950. Следовательно, среди 247 480 работ, цитирующих статью [11] и, соответственно, использующих идеи О. Фолина, на которых она основана, почти нет таких, которые явно ссылаются на его статью [10].

Цель настоящей статьи состоит в модификации индекса цитирования за счет разработки механизмов компенсации указанных выше недостатков. Идея заключается в дополнении индекса цитирования составляющей, учитывающей диффузию знаний. В [12] индикатором диффузии знаний выбра-

но цитирование работы. Но это прямое, видимо, использование знаний. Существует и скрытая диффузия знаний, состоящая в использовании идей без прямой ссылки на ее автора, но с возможностью идентификации первоисточника через цепочку цитирований. В приведенном выше примере это цитирование только работы А. Лоури без прямой ссылки на статью О. Фолина. Такие случаи назовем скрытой диффузией знаний, а цепочку ссылок, приводящую к первоисточнику, назовем неявным цитированием. Соответственно, подсчитав каким-то образом число неявных цитирований и добавив его с некоторым весом к обычному индексу цитирования, получим рейтинг ученого с учетом скрытой диффузии знаний. Аналогичные принципы оценивания персонала используются в сетевом маркетинге, когда сотрудник получает баллы как за свои прямые продажи, так и за продажи, которые осуществила завербованная им команда. Здесь персональные продажи сотрудника соответствуют прямому цитированию, а продажи его команды – неявному цитированию. Эти же принципы применяются в спорте для оценки по системе "гол плюс пас" атакующих игроков команды. Соответственно, аналогом гола является прямое цитирование, а паса – неявное.

## ФОРМАЛИЗАЦИЯ ПОСТАНОВКИ ЗАДАЧИ

Будем считать известным список из  $M$  публикаций в таком формате:

$$\mathbf{A}_j = \langle j, \mathbf{F}_j, \mathbf{T}_j \rangle, \quad j = \overline{1, M}, \quad (1)$$

где  $\mathbf{F}_j = \{F_j^1, F_j^2, F_j^3, \dots\}$  – множество номеров публикаций, на которые ссылается работа  $\mathbf{A}_j$ ;

$\mathbf{T}_j = \{T_j^1, T_j^2, T_j^3, \dots\}$  – множество номеров публикаций, в которых цитируется работа  $\mathbf{A}_j$ .

С математической точки зрения, задача состоит в нахождении отображения  $\mathbf{A}_j \rightarrow C_j$  ( $j = \overline{1, M}$ ), где  $C_j$  – новый индекс цитирования, который учитывает скрытую диффузию знаний.

## УЧЕТ СКРЫТОЙ ДИФфуЗИИ ЗНАНИЙ

Индекс цитирования, учитывающий скрытую диффузию знаний, определим следующим образом:

$$C_j = D_j + \alpha \cdot I_j, \quad (2)$$

где  $D_j = |\mathbf{T}_j|$  – обычный индекс цитирования, учитывающий только прямые ссылки ( $|\cdot|$  – обо-

значает мощность множества, т.е. количество его элементов);

$I_j$  – индекс неявного цитирования;

$\alpha \in [0, 1]$  – весовой коэффициент важности неявного цитирования.

В случае  $\alpha = 0$  новый показатель (2) эквивалентен обычному индексу цитирования. Когда  $\alpha = 1$  вклады прямого и скрытого цитирования будут равнозначными.

Диффузию знаний учтем следующим образом. Каждая научная работа содержит как новые знания, так и знания из цитируемых работ. Предположим, что доля заимствованных знаний постоянна. Обозначим ее через  $\beta \in (0, 1)$ . Рассмотрим простую цепочку цитирований  $C \rightarrow B \rightarrow A$ , когда работа  $A$  ссылается только на работу  $B$ , а работа  $B$  – только на работу  $C$ . Соответственно,  $\beta$ -доля знаний перешла из  $B$  в  $C$ , а затем  $\beta$ -доля знаний из  $B$  перешла в  $A$ . Тогда, из  $C$  в  $A$  перешли знания, доля которых составляет  $\alpha = \beta \cdot \beta$ , хотя это и не подтверждается прямым цитированием. Такое неявное использование знаний или скрытая диффузия обусловлена тем, что автор включает в перечень литературы наиболее релевантные источники, уровень использования которых превышает некоторый порог [8].

Рассмотрим случай, когда список литературы в  $B$  состоит из  $N$  источников, среди которых и работа  $C$ . Тогда будем считать, что из  $C$  в  $B$  перешли знания, доля которых составляет  $\frac{\beta}{N}$ . Обобщая выкладки, получаем такую формулу индекса неявного цитирования:

$$I_j = \sum_{\substack{\forall i \in \mathbf{T}_j \\ j \notin \mathbf{F}_i}} \frac{|\mathbf{T}_i|}{N_i}, \quad (3)$$

где  $N_i = |\mathbf{F}_i|$  – длина списка литературы в  $i$ -й публикации.

Условие  $j \notin \mathbf{F}_i$  в (3) указывает, что в случае прямого цитирования неявное цитирование не учитывается.

**Пример.** Исходные данные приведены в табл. 1. Соответствующая сеть цитирования изображена на рисунке. Расчет нового индекса цитирования (2) по этим данным иллюстрирует табл. 2. Из нее видно, что публикации № 2, 4, 5 и 6 получили одинаковое количество прямых цитирований. Среди них самый высокий рейтинг имеет публикация №4 в связи с максимальным количеством неявных цитирований, две из которых посредством работы № 6 и пять – посредством работы № 7.

Таблица 1

Исходные данные по цитированию в формате (1)

$j$	$F_j$	$T_j$
1	неизвестно	{8}
2	неизвестно	{6, 8}
3	неизвестно	{6}
4	неизвестно	{6, 7}
5	неизвестно	{7, 12}
6	{2, 3, 4}	{8, 9}
7	{4, 5}	{8, 9, 10, 11, 12}
8	{1, 2, 6, 7}	–
9	{6, 7}	–
10	{7}	–
11	{7}	–
12	{5, 7}	–

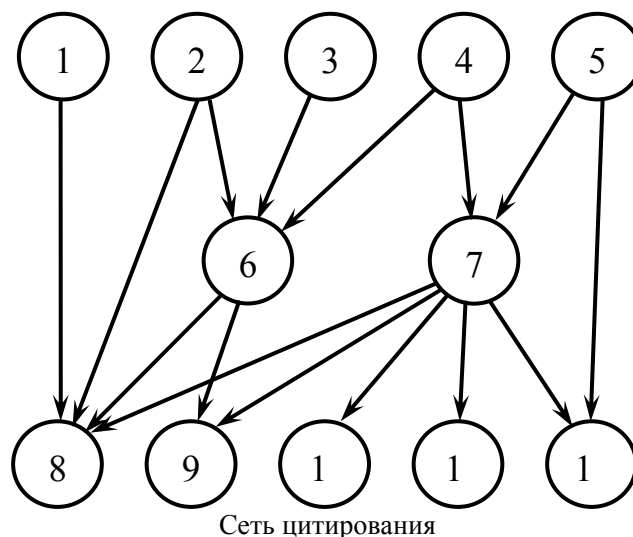


Таблица 2

К расчету индексу цитирования (2)

$j$	1	2	3	4	5	6	7
$D_j$	1	2	1	2	2	2	5
$I_j$	0	$\frac{1}{3} = 0,33$	$\frac{2}{3} = 0,67$	$\frac{2}{3} + \frac{5}{2} = 3,17$	$\frac{4}{2} = 2$	0	0
$C_j, \alpha = 0,1$	1	2,03	1,07	2,32	2,2	2	5
$C_j, \alpha = 0,5$	1	2,17	1,33	3,58	3	2	5

## СРАВНЕНИЕ С АНАЛОГАМИ

Нами обнаружено, что для ранжирования научных публикаций неявное цитирование используется только в [13] для расчета индекса Хирша отдельной статьи. В этом случае, индекса Хирша статьи составляет  $h$ , если на нее ссылаются  $h$  работ, каждую из которых процитирован минимум  $h$  раз.

Ближайшим аналогом является подход [14], в котором неявные ссылки используются для расчета кумулятивного индекса цитирования патентов. Для этого применяется аналог формулы (2) для всей цепочки цитирований. При этом, чем длиннее цепочка неявного цитирования, тем меньше весовой коэффициент. Принципиальное отличие нашего подхода заключается в учете длины списка литературы при расчете неявного цитирования в (3), тогда как в [14] длина соответствующего списка патентов не учитывается. Игнорирование длины списка литературы может привести к тому, что составляющая от неявного цитирования будет больше, чем если бы имело место прямое цитирование. Например, для фрагмента сети цитирований с вершинами 4, 6, 7 и 8 (см. рис. 1), работа № 8 обеспечила бы 2 балла работе № 4 за неявное цитирование, тогда как в случае прямого цитирования  $4 \rightarrow 8$  работа №4 получила лишь бы 1 балл.

## ВЫВОДЫ

Предложен новый наукометрический показатель, оценивающий диффузию знаний. Он имеет две составляющие, первая из которых эквивалентна обычному индексу цитирования, т. е. описывает видимую диффузию научных знаний. Вторая составляющая отражает скрытую диффузию научных знаний и выражается через количество неявных цитирований. Предложенный подход наследует механизмы оценки деятельности персонала в сетевом маркетинге и атакующих игроков в хоккее, футболе и других командных играх спорта.

Практическая ценность предложенного показателя заключается в том, что он позволяет легко идентифицировать креативных ученых, которые сгенерировали новые идеи сподвигнули научное сообщество на создание значительного количества востребованных (высокоцитированных) работ. При этом, сами генераторы идей остались в тени. За такими учеными часто «охотятся» лидерские исследовательские компании, но автоматически их обнаружить было достаточно сложно из-за низких наукометрических показателей. Теперь идентификацию скрытых инициаторов научного мейнстрима можно формализовать, выявив ученых, которые одновременно имеют большое значение предложенного индекса цитирования и малое значение традиционного индекса цитирования. Практическое применение

предложенного показателя возможно лишь на основе автоматизированного учета научных публикаций, например, на базе систем Web of Knowledge, Scopus, Google Scholar или eLibrary.ru.

#### СПИСОК ЛИТЕРАТУРЫ

1. Egghe L. Mathematical theory of the h- and g-index in case of fractional counting of authorship // Journal of the American Society for Information Science and Technology. – 2008. – Vol. 59, № 10. – P. 1608–1616.
2. 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.
3. Buena-Casal G. Assessing the quality of articles and scientific journals: Proposal for weighted impact factor // Psychology in Spain. – 2004. – Vol. 8, № 1. – P. 60–76.
4. Van Noorden R. A profusion of measures // Nature. – 2010. – Vol. 465. – P. 864–866.
5. Garcia-Perez M.A. The Hirsch h index in a non-mainstream area: methodology of the behavioral sciences in Spain // The Spanish Journal of Psychology. – 2009. – Vol. 12, № 2. – P. 833–849.
6. Сойфер В.Н. Международная Соросовская программа образования. Часть 2. Результаты именных конкурсов // Соросовский образовательный журнал. – 1996. – № 1. – С. 4–16.
7. Franceschini F., Maisano D., Perotti A., Proto A. Analysis of the ch-index: an indicator to evaluate the diffusion of scientific research output by citers // Scientometrics. – 2010. – Vol. 85. – P. 203–217.
8. Кара-Мурза С.Г. Цитирование в науке и подходы к оценке научного вклада // Вестник АН СССР. – 1981. – № 5. – С.68-75.
9. Chen C., Hicks D. Tracing knowledge diffusion // Scientometrics. – 2004. – Vol. 59, № 2. – P. 199–211.
10. Folin O., Ciocalteu V. On tyrosine and tryptophane determinations in proteins // The Journal of Biological Chemistry. – 1927. – Vol. 73, № 2. – P. 627–650.
11. Lowry O.H., Rosbrough N.J., Farr A.L., Randall R.J. Protein measurement with the Folin phenol reagent // The Journal of Biological Chemistry. – 1951. – Vol. 193, № 1. – P. 265–275.
12. Gao X., Guan J. Network model of knowledge diffusion // Scientometrics. – 2012. – Vol. 90, № 3. – P. 749–762.
13. Schubert A. Using the h-index for assessing single publications // Scientometrics. – 2009. – Vol. 78, № 3. – P. 559–565.
14. Atallah G., Rodriguez G. Indirect patent citations // Scientometrics. – 2006. – Vol. 67, № 3. – P. 437–465.

*Материал поступил в редакцию 02.04.13.*

#### Сведения об авторах

**ШТОВБА Сергей Дмитриевич** - доктор технических наук, профессор кафедры компьютерных систем управления Винницкого национального технического университета, Украина  
E-mail: shtovba@gmail.com

**ШТОВБА Елена Валерьевна** - кандидат экономических наук, доцент кафедры менеджмента и моделирования в экономике, Винницкого национального технического университета, Украина  
E-mail: olena.shtovba@yahoo.com