

С. Д. Штовба, д-р. техн. наук, проф.; О. В. Штовба, канд. екон. наук

ЗАСТОСУВАННЯ ПРИНЦИПІВ СІТЬОВОГО МАРКЕТИНГУ ДЛЯ ОЦІНЮВАННЯ ДИФУЗІЇ НАУКОВИХ ЗНАНЬ

Запропоновано новий наукометричний показник, що оцінює дифузію знань. Він має дві складові, які визначаються за кількістю прямих та прихованих цитувань. Запропонований підхід успадковує механізми оцінювання діяльності персоналу в сітьовому маркетингу. Аналогом прямого цитування виступають персональні продажі працівника, а непрямого цитування — продажі його команди.

Вступ

Сьогодні для оцінювання корисності дослідницької діяльності спільно з експертними оцінками частіше використовують і наукометричні показники. В першу чергу наукометричні показники використовують для оцінювання фундаментальних досліджень, результати яких безпосередньо не пов'язані з економічним ефектом. Результати фундаментальних досліджень спрямовано на розвиток науки, тому їх важливість оцінюють через відклик наукової спільноти. Формально, цей відклик виражають індексом цитування — сумарною кількістю публікацій з посиланням на аналізовані роботи. Зазвичай, слабкі роботи не цитують, за винятком особливих стосунків між авторами. Основні варіанти індексу цитування модифікують його таким чином:

- 1) не враховують самоцитування або цитування співавторами [1], що суттєво знижує рейтинг «науковця-затворника», публікації якого цікавлять лише його самого;
- 2) не враховують повторне цитування однієї роботи одним і тим же науковцем [2], що зменшує вплив компліментарного цитування;
- 3) враховують особистий внесок науковця [3, 4], розділяючи кількість цитувань роботи на кількість її співавторів, яких в одній статті сьогодні може бути більше 3000;
- 4) враховують авторитетність видання, в якому цитується робота аналізованого науковця, шляхом множення кількості цитувань на відповідний коефіцієнт, наприклад, на імпаکت-фактор журналу [5, 6].

Але використання цих та інших модифікацій не усувають таких двох недоліків індексу цитування. Перший недолік пов'язаний із явищем забування імен класиків, коли автори вважають, що внесок попередників настільки добре відомий будь-кому із цієї галузі науки, що немає сенсу про це згадувати [7]. Відповідно, роботи таких науковців вищого рангу швидко перестають цитувати. Другий недолік індексу цитування полягає у низькому рейтингу концептуальних робіт, що описують нові ідеї. Це відбувається через достатньо поширене явище приховування першоджерел, тобто невключення концептуальних робіт попередників до переліку цитованої літератури. Створення нового наукового результату — це багатоетапний процес, що включає [7]: 1) генерування ідеї та розробку концепції; 2) її модифікації для підвищення точності, швидкодії, спрощення тощо; 3) просування та практичне використання. В історії науки є багато фактів, коли статті, що стосуються другого та третього етапів, отримують значно більше цитувань, ніж концептуальні роботи. Найяскравішим прикладом, серед наведених у [7], є статі О. Фоліна [8] та О. Лоурі [9]. О. Лоурі модифікував реактив О. Фоліна із колориметричного визначення білка шляхом додавання ще одного компонента, що розширило границі застосування методу. Сьогодні, за даними Google Scholar статтю О. Лоурі процитовано 228 295 разів, тоді як роботу О. Фоліна — 1611. Тобто, серед 228 295 робіт, що цитують статтю [9] і, відповідно, використовують ідеї О. Фоліна, на яких вона ґрунтується, майже немає таких, що явно посилаються на його статтю [8].

Метою статті є модифікація індексу цитування, за рахунок розробки механізмів компенсації вказаних вище недоліків. Ідея полягає у доповненні індексу цитування складовою, що враховує дифузію знань. В [10] індикатором дифузії знань обрано цитування роботи. Але це пряме — видиме використання знань. Існує і прихована дифузія знань, що полягає у викорис-

танні ідей без прямого посилання на її автора, але з можливістю ідентифікації першоджерел через ланцюжок цитувань. В наведеному вище прикладі це цитування лише роботи О. Лоурі без прямого посилання на статтю О. Фоліна. Такі випадки назвемо прихованою дифузійною знань, а ланцюжок цитувань, що веде до першоджерела назвемо непрямим цитуванням. Відповідно, підрахувавши у якійсь спосіб число непрямих цитувань та додавши його з деякою вагою до звичайного індексу цитування отримуємо рейтинг науковця з урахуванням прихованої дифузії знань. Схожі механізми оцінювання персоналу діють у сітьовому маркетингу, коли працівник отримує бали як за свої прямі продажі, так і бали з продажів, які здійснили завербовані ним працівники нижчого рівня ієрархії. Тут прямому цитуванню відповідають персональні продажі працівника, а непрямому цитуванню – продажі його команди завербованих працівників. Аналогічні механізми використовують і у спорті для оцінювання гравців команди, що атакують, за системою «гол плюс пас». Відповідно, аналогом голу є пряме цитування, а пасу – непряме.

Математична постановка задачі

Нехай відомим є перелік із M публікацій у такому форматі:

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

де $\mathbf{F}_j = \{F_j^1, F_j^2, F_j^3, \dots\}$ – множина номерів публікацій, які цитуються в роботі \mathbf{A}_j ;

$\mathbf{T}_j = \{T_j^1, T_j^2, T_j^3, \dots\}$ – множина номерів публікацій, які цитують роботу \mathbf{A}_j .

Задача полягає у знаходженні відображення $\mathbf{A}_j \rightarrow C_j$, де C_j – новий індекс цитувань, що враховує приховану дифузійною знань.

Індекс цитувань за прихованою дифузійною знань

Індекс цитувань, що враховує приховану дифузійною знань, задамо таким чином:

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

де $D_j = |\mathbf{T}_j|$ – звичайний індекс цитувань, що враховує прямі посилання ($|\cdot|$ – означає потужність множини); I_j – індекс непрямих цитувань; $\alpha \in [0, 1]$ – ваговий коефіцієнт важливості непрямих цитувань.

У випадку $\alpha = 0$, новий індекс (2) тотожний звичайному індексу цитувань. У випадку $\alpha = 1$ – складові за прямим та прихованим цитуванням рівноважні.

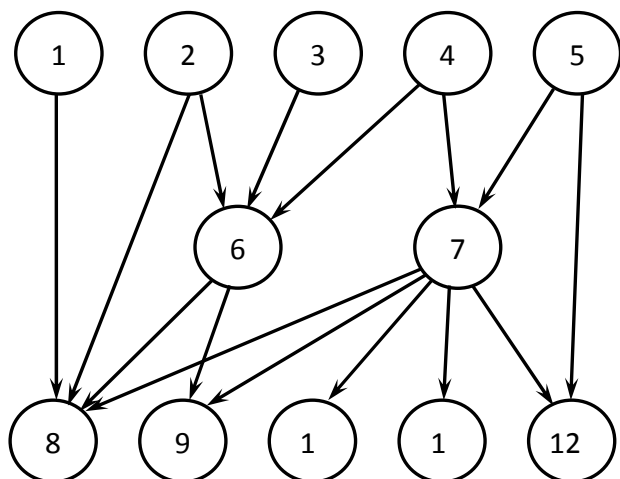
Дифузійною знань врахуємо у такий спосіб. Вважатимемо, що кожна наукова робота містить нові знання та запозичені знання з цитованих робіт. Вважатимемо частку запозичених знань сталою та позначимо її через $\beta \in (0, 1)$. Розглянемо простий ланцюжок цитувань $C \rightarrow B \rightarrow A$, коли робота A посилається лише на роботу B , а робота B – лише на роботу C . Відповідно, β -частка знань перейшла з B в C , та β -частка знань перейшла з B в A . Тоді, частка знань у обсязі $\alpha = \beta \cdot \beta$ перейшла з C в A , хоча це і не засвідчується прямим цитуванням. Такий непрямий перехід знань або прихована дифузійною знань обумовлена тим, що автор включає у перелік літератури найбільш релевантні джерела, рівень використання яких перевищує деяких поріг [7]. У випадку, коли перелік літератури в B складається із n джерел серед яких і робота C , тоді вважатимемо, що з C в B перейшла частка знань у обсязі $\frac{\beta}{N}$. Узагальнюючи ці викладки, отримуємо таку модель індексу непрямих цитувань:

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

де $N_i = |\mathbf{F}_i|$ – довжина списку літератури в i -й публікації.

Умова $j \notin F_i$ в (3) вказує, що непрямі цитування не розраховуються, якщо має місце пряме цитування.

Приклад. Початкові дані наведено в табл. 1. Відповідна мережа цитувань зображена на рисунку. Розрахунок індексу цитувань (2) за цими даними ілюструє табл. 2. З неї видно, що публікацій №№ 2, 4, 5 та 6 отримали однакову кількість прямих цитувань. Серед них найвищий рейтинг має публікація 4 завдяки максимальній кількості непрямих цитувань — 2 через роботу 6 та 5 через роботу 7.



Таблиця 1
Початкові дані із цитування у форматі (1)

j	F_j	T_j
1	невідомо	{8}
2	невідомо	{6, 8}
3	невідомо	{6}
4	невідомо	{6, 7}
5	невідомо	{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}	–

Таблиця 2

До розрахунку індексу цитувань (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

Порівняння з найближчими аналогами

Авторами виявлено, що для ранжування наукових публікацій непрямі цитування використовуються лише в [11] для розрахунку індексу Хірша окремої статті. В цьому випадку індекс Хірша статті дорівнює h , якщо на неї посилається h робіт, кожену із яких процитовано щонайменше h раз.

Ближчим аналогом є стаття [12], в якій непрямі посилання використовуються для визначення кумулятивного індексу цитувань патентів. Для цього застосовується аналог формули (2) для усього ланцюжка цитувань. При цьому чим довший ланцюжок непрямого цитування, тим менше вагомий коефіцієнт α . Принципова відмінність нашого підходу полягає у врахуванні довжини списку літератури під час розрахунку непрямого цитування в (3), тоді як в [12] довжина відповідного списку патентів до уваги не береться. Не врахування довжини списку літератури може привести до того, що складова від непрямого цитування буде більша ніж, якби мало місце пряме цитування. Наприклад, для фрагмента мережі цитувань з вершинами 4, 6, 7 та 8 (див. рис. 1), робота № 8 забезпечила би 2 бали роботі № 4 за непряме цитування, тоді як за прямого цитування $4 \rightarrow 8$ робота № 4 отримала лише б 1 бал.

Висновки

Запропоновано новий наукометричний показник, що оцінює дифузію знань. Він має дві складові, перша з яких еквівалентна звичайному індексу цитувань, тобто описує видиму дифузію наукових знань. Друга складова віддзеркалює приховану дифузію наукових знань і залежить від кількості непрямих цитувань. Запропонований підхід успадковує механізми оцінювання діяльності персоналу в сітьовому маркетингу, ставляючи у відповідність персональним продажам працівника пряме цитування, а продажам його команди — непряме цитування.

Практична цінність запропонованого показника полягає в тому, що за ним просто ідентифікувати креативних науковців, які згенерували нові ідеї, що спонукали наукову спільноту на створення значної кількості високоцитованих робіт. При цьому самі генератори ідей залишилися в тіні. За такими науковцями часто «полюють» лідерські дослідницькі компанії, але автоматично виявити їх було достатньо складно через низький індекс цитування. Тепер це можна зробити у формальний спосіб, визначивши науковців, що одночасно мають велике значення запропонованого індексу цитування та мале значення традиційного індексу цитування. Практичне застосування запропонованого показника можливе лише на основі автоматизованого обліку наукових публікацій, наприклад, на базі систем Web of Knowledge, Scopus, Google Scholar чи eLibrary.ru.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Сойфер В. Н. Международная Соросовская программа образования. Часть 2. Результаты именных конкурсов / В. Н. Сойфер // Соросовский образовательный журнал. — 1996. — № 1. — С. 4—16.
2. Analysis of the ch-index: an indicator to evaluate the diffusion of scientific research output by citers / [Franceschini F., Maisano D., Perotti A., Proto A.] // *Scientometrics*. — 2010. — Vol. 85. — P. 203—217.
3. Egghe L. Mathematical theory of the h- and g-index in case of fractional counting of authorship / L. Egghe // *Journal of the American Society for Information Science and Technology*. — 2008. — Vol. 59, № 10. — P. 1608—1616.
4. Schreiber M. A modification of the h-index: The h(m)-index accounts for multi-authored manuscripts / Schreiber M. // *Journal of Informetrics*. — 2008. — Vol. 2, № 3. — P. 211—216.
5. Buela-Casal G. Assessing the quality of articles and scientific journals: Proposal for weighted impact factor / G. Buela-Casal // *Psychology in Spain*. — 2004. — Vol. 8, № 1. — P. 60—76.
6. Van Noorden R. A profusion of measures / Van Noorden R. // *Nature*. — 2010. — Vol. 465. — P. 864—866.
7. Кара-Мурза С. Г. Цитирование в науке и подходы к оценке научного вклада / С. Г. Кара-Мурза // *Вестник АН СССР*. — 1981. — № 5. — С. 68—75.
8. Folin O. On tyrosine and tryptophane determinations in proteins / O. Folin, V. Ciocalteu // *The Journal of Biological Chemistry*. — 1927. — Vol. 73, № 2. — P. 627—650.
9. Protein measurement with the Folin phenol reagent / O. H. Lowry, N. J. Rosbrough, A. L. Farr, R. J. Randall // *The Journal of Biological Chemistry*. — 1951. — Vol. 193, № 1. — P. 265—275.
10. Gao X. Network model of knowledge diffusion / X. Gao, Guan J. // *Scientometrics*. — 2012. — Vol. 90, № 3. — P. 749—762.
11. Schubert A. Using the h-index for assessing single publications / Schubert A. // *Scientometrics*. — 2009. — Vol. 78, № 3. — P. 559—565.
12. Atallah G. Indirect patent citations / G. Atallah, G. Rodriguez // *Scientometrics*. — 2006. — Vol. 67, № 3. — P. 437—465.

Рекомендована кафедрою менеджменту та моделювання в економіці

Стаття надійшла до редакції 13.03.12
Рекомендована до друку 3.04.12

Штовба Сергій Дмитрович — професор кафедри комп'ютерних систем управління;

Штовба Олена Валеріївна — доцент кафедри менеджменту та моделювання в економіці.

Вінницький національний технічний університет, Вінниця