УДК 002 (477):04.67

S. D. Shtovba, Dc. Sc. (Eng.), Assoc. Prof.; O. V. Shtovba, Cand. Sc. (Economics)

SH-INDEX – A NEW FRACTIONAL MODIFICATION OF HIRSCH-INDEX

A new scientometric indicator – Sh-index is proposed. In Sh-index the integer part is a usual Hirsch-index and the fractional part shows how close a scientist is to the next value of Hirsch-index. The fractional part is calculated as a proportion of published papers required to reach the next value of Hirsch-index.

Key words: *scientometrics, citation index, Hirsch-index (h-index), fractional modification, Sh-index.*

Introduction

Today, in order to evaluate the productivity of research activities, along with expert opinions scientometric indices are being increasingly used. The interest to scientometric indices was catalyzed by the works of V.V. Nalimov, a Soviet scientist. The most famous among them is monograph [1]. Initially, the main scientometric indicator was the number of published works of a scientist – the total number or of a certain publication types: monographs, papers, theses, publications in specialized editions included into the list of Higher Attestation Commission, into electronic databases Web of Science or Scopus, indexed by Google Scholar etc. For artificial improvement of this indicator the following typical techniques are used: division of the results in order to publish them in different editions or even duplication of the description of the same materials in the articles having different titles. Thus, pursuit of the number of publications often reduces quality of scientific works.

The citation index – the number of publications with the references to the author whose works are being analyzed – has become the second scientometric indicator. Citation index reflects reaction of the scientific community to the author's publications and, therefore, is feasible to be used for evaluation of fundamental research results. As a rule, insignificant publications are not cited except in the cases of special relations between the authors. Different modifications of the citation index clear it from "noise", e. g. exclude self-citations and citations by coauthors or require both the presence of the paper and of the works with references to this paper in the same database. In order to get into this database the papers should meet definite requirements. However, this index also does not protect from fragmentation of the research results and publishing them in smaller papers. High citation index could be provided by a single paper that the author wrote many years ago and had not published anything significant since that time.

In 2005, to identify the scholars who write much and qualitatively, J. Hirsch, a physicist, proposed a new indicator – the index of Hirsch [2]. Hirsch index, or h-index, is a maximal integer number h indicating that the author published h papers, each of them being cited at least h times. These h papers form the core of Hirsch. To get into Hirsch core the paper is to be cited at least h times and so the results should not be divided for publishing in several papers. The simplicity of calculations and insensitiveness to the typical techniques for artificial improvement of the indices, considered above, instantly made Hirsch index a popular scientometric indicator.

Let us consider two authors who with minimal permissible requirements received neighboring values of Hirsch-index – h and $h+1$. The number of citations of these authors differs significantly, and namely, by the value $(h+1)^2 - h^2 = 2h+1$. Therefore, in accordance with Hirsch-index the authors could be easily clustered into groups with its equal values. However, integrality of this indicator does not make it possible to rank authors within a group with the same values of Hirsch-index. **The aim of the paper** is to develop a new scientometric indicator that, inheriting all the advantages of Hirsch-index, will additionally enable such ranking. The new indicator is a fractional

modification of Hirsch-index. Its integer part is an ordinary Hirsch index and a fractional part shows how close the author approached the next value of h-index. In accordance with the traditions of scientometrics, the new indicator will be called Sh-index.

A formalized statement of the problem

Let us designate the number of publications of a certain author as N. The number of citations of this author is sorted in descending order and presented by the following vector:

$$C = (c_1, c_2, ..., c_N), \tag{1}$$

where c_i is the number of citations of the i-th publication and $c_1 \geq c_2 \geq ... \geq c_N$, $i = \overline{1, N}$.

The objective of the research is to find a functional mapping

$$C \to Sh\_index \in [h(C), h(C)+1], \tag{2}$$

where Sh_index is fractional modification of Hircsh-index (Sh-index);

$h(C) = \max\limits_{\substack{i=1, N \\ i \geq c_i}}(i)$ – Hirsch-index.

On the map (2) the following restrictions are imposed:

(a) if Hirsch-index $h(C)$ is achieved with minimum possible number of citations, i.e. in the case of $(c_1 = h, c_2 = h, ..., c_h = h)$, then $Sh\_index(C) = h(C)$;

(b) the greater the fractional part of Sh_index, the closer is the author to the increase of his Hirsch-index.

The new Sh-индекс

The new scientometric Sh-index for the vector of citations (1) we define as follows:

$$Sh\_index(C) = h(C) + \Delta, \tag{3}$$

where $\Delta = \frac{1}{h+1} \cdot \max\limits_{\substack{j=1, h \\ c_j > h}}(j)$ is a fractional part of Sh-index indicating the proportion of fullness of the current h-core with the papers that, by the number of citations, are also included into the next core.

For the convenience of calculations we rewrite the fractional component (30 in the following form:

$$\Delta = \frac{\sum\limits_{j=1, h} r_j}{h+1}, \tag{4}$$

where $r_j = \begin{cases} 1, & if\ c_j > h \\ 0, & otherwise \end{cases}$ is the indicator of the paper membership to the next Hirsch core.

Example. Citation vectors of three authors are presented in table 1. All the authors have the same Hirsch index h=4. In order to increase it by 1, 5 papers are required and each of them must be cited at least 5 times. Author A has 4 such papers, author B – 3 and author C – 0 such papers. Hence, the estimates of the authors in accordance with the proposed Sh-index are as follows:

$sh\_index(Author\ A) = 4 + \frac{4}{5} = 4.8$;

$sh\_index(AuthorB) = 4 + \frac{3}{5} = 4.6$;

$sh\_index(Author\ C) = 4 + \frac{0}{5} = 4$.

Table 1

Vectors of citation

Authors	Number of the paper							
	1	2	3	4	5	6	7	8
Author A	20	9	7	5	4	2	1	0
Author B	6	5	5	4	1	0	0	0
Author C	4	4	4	4	0	0	0	0

Comparison with competitive approaches

For ranking scientists having identical Hirsch-indices additional indicators are mostly used. The most popular among them are summarized in table 2. An advantage of the proposed index consists in that it generalizes Hirsch-index while indices from table 2 are used in conjunction with it. Advantages and disadvantages of those additional scientometric indicators are analyzed in many papers, e.g. in [9, 10] and, therefore, we will not discuss them here.

Table 2

Additional scientometric indices

Indicator	Description	Relations	The source
g-index	Maximal number g of the most popular papers that received in total no less than g^2 citations. It takes into account how the summarized citing of Hirsch core exceeds minimal requirements.	$h \le g$	[3]
hg-index	Mean geometric value of h-index and g-index: $hg = \sqrt{h \cdot g}$.	$h \le hg \le g$	[4]
e-index	Square root of the excessive citing of Hirsch core h^2 : $e = \sqrt{\sum_{j=\overline{1,h}} c_j - h^2}$.	–	[5]
A-index	The average number of citations of Hirsch core: $A = \frac{1}{h} \sum_{j=\overline{1,h}} c_j$.	$R = \sqrt{h \cdot A}$	[6]
R-index	Square root of the summarized citing of Hirsch core : $R = \sqrt{\sum_{j=\overline{1,h}} c_j}$.	$R = \sqrt{h \cdot A}$	[7]
m-index	Median of the number of citations of Hirsch core.	$h \le m$	[8]

The closest analog of the proposed Sh-index is h_{rat}-index [11]:

$$h_{rat}(C) = h(C) + 1 - \frac{n}{2h(C)+1}, \qquad (5)$$

where $n = (h(C) + 1 - c_{j+1}) + \sum_{j=\overline{1,h}} (1 - r_j)$ – minimal number of citations required for approaching the next value of Hirsch index.

$2h(C)+1$ is theoretically minimal number of citations required for reaching the next value of Hirsch index $(h+1)$ in the case of the worst initial conditions when $(c_1 = h, c_2 = h, ..., c_h = h, c_{h+1} = 0)$.

Comparing (3) and (5), we see that integer parts of both indices are identical and correspond to Hirsch index. The fractional parts differ due to the different principles of interpolation of the segment $[h, h+1]$. In (5) this unit segment is divided into $2h(C)+1$ equal parts in accordance with the minimal required number of citations. Then to h number of publications from the current Hirsch core the proportion of implementing the citation plan on the next Hirsch core formation is added. Hence, in h_{rat} there is a contradiction between the components because one of the summands corresponds to the number of publications and the second is formed according to the number of

citations. In the proposed approach this shortcoming is eliminated as both the integer and the fractional parts of Sh-index (3) are composed according to the number of papers. In (3) the integer part is equal to the number of papers from the current Hirsch core and the fractional part – to the proportion of the next core fullness with papers. In other words, the fractional part of the Hirsch-index could be interpreted as a proportion of implementing the plan on papers in order to obtain the next value of Hirsch-index.

Conclusions

A new scientometric indicator – Sh-index – is proposed. This index is a fractional modification of one of the most popular scientometric indicators – the Hirsch index. In Sh-index the integer part is a usual Hirsch-index and the fractional component shows how close the author is to the next value of Hirsch-index. The fractional part is calculated as a proportion of the already published papers that are required to reach the next value of Hirsch-index. Inheriting all the advantages of Hirsch-index, the new indicator additionally enables to rank the authors with equal Hirsch indices. Besides, this new indicator makes it possible to trace the progress of a scientist in improving his Hirsch-index within a definite period of time.

REFERENCES

1. Налимов В. В. Наукометрия. Изучения развития науки как информационного процесса / В. В. Налимов, З. М. Мульченко. – М.: Наука. – 1969. – 192 с.

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

3. Egghe L. Theory and practice of the g-index / L. Egghe // Scientometrics. – 2006. – Vol. 69, № 1. – P. 131 – 152.

4. Alonso S. Hg-index: a new index to characterize the scientific output of researchers based on the hand g-indices / S. Alonso, F. Cabrerizo, E. Herrera-Viedma, F.Herrera // Scientometrics. – 2010. – Vol. 82, № 2. – P. 391 – 400.

5. The e-index, complementing the h-index for excess citations [Електронний ресурс] / Zhang C.-T. // PLoS ONE. – Vol. 4, №5. – 2009. Режим доступу:
http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0005429 .

6. New developments related to the Hirsch index [Електронний ресурс] / Rousseau R. // Science Focus. – Vol. 1, № 4. – 2006. – P. 23 – 25. Режим доступу: http://eprints.rclis.org/6376/.

7. Liang B. J. L. The R- and AR-indices: complementing the h-index / B. J. L. Liang, R. Rousseau , L. Egghe // Chinese Science Bulletin. – 2007. – Vol. 52, № 6. – P. 855 – 863.

8. Bornmann L. Are there better indices for evaluation purposes than the h-index? A comparison of nine different variants of the h-index using data from biomedicine / L. Bornmann, R. Mutz, H.Daniel // Journal of the American Society for Information Science and Technology. – 2008. – Vol. 59, № 5. – P. 830 – 837.

9. Alonso S. H-index: a review focused in its variants, computation and standardization for different scientific fields / S. Alonso, F. J. Cabrerizo, E. Herrera-Viedma, F. Herrera // Journal of Informetrics. – 2009. – Vol. 3, № 4. – P. 273 – 289.

10. Egghe L. The Hirsch-index and related impact measures / L. Egghe // Annual Review of Information Science and Technology. – 2010. – Vol. 44. – P. 65 – 114.

11. Guns R. Real and rational variants of the h-index and the g-index / R. Guns, R. Rousseau // Journal of Informetrics. – 2009. – Vol 3, № 1 – P. 64 – 71.

Shtovba Sergiy – Prof. of the Department of Computer Control Systems.

Shtovba Olena – Assoc. Prof. of the Department of Management and Modeling in Economics.
Vinnytsia National Technical University.