C-Index as a Quantitative Index to Measure Individual Academic Achievement in China

Cai Ning[a] and Fu Yuanyuan*[b]

College of Electrical Engineering, Northwest University for Nationalities, Lanzhou, China

[a]caining91@tsinghua.org.cn, [b]785325020@qq.com

Keywords: Index; Scientific achievement; h-index; Times of cites; Document set.

Abstract: This paper proposes a new index for evaluating the scientific achievement of individual scientists, based on improving the popular h-index, whilst endeavoring to keep the merits of h-index and remove the demerits. The gist of computation algorithm for the new index is to regard the average times of cites for the publications selected. Besides, several accessional principles are also discussed, with considering the specific situation of China.

Introduction

Nowadays, h-index has become a standard to quantitatively evaluate academic achievement. h-index is a mixture of quantitative indicators, proposed by physicist George Hirsch (Jorge Hirsch) with the University of California, San Diego, in 2005 [1]. Its aim is to quantitatively measure the academic achievement of certain individual researcher. Concretely speaking, the definition of the h-index is the number of h of a scientist, and a scientist has index h if h of his/her N papers have at least h citations each, with the other ($N-h$) papers possessing no more than h citations each. So far, h-index has been widely used to evaluate the level of scholarships in all areas of academics [2-3]. For instance, h-index becomes the referential basis for different levels of academic titles. In USA, usually any tenured associate professor in the natural sciences of any university should have an h-index not less than 10. A survey by the London School of Economics [4] reveals that (full) professors in the social sciences have average h-indices ranging from 2.8 (in law) to 7.6 (in economics). On average across the disciplines, a full professor in the social sciences has an h-index of 4.9, whereas a senior lecturer (associate professor) has an h-index of 2.2. Moreover, this index can also be applied to the productivity and impact of a scholarly journal as well as a group of scientists, such as a department or university or country [5].

Nonetheless, h-index is far from being perfect, especially for China. Criticisms have been kept on emerging toward h-index [6-9]. From our point of view, h-index has two significant drawbacks: 1) It regards the article with lowest number of citations and neglects those highly cited; 2) it indiscriminately treats all the authors of an article, and can thereby be easily operated by deliberate trickery [10]. Due to these reasons, in many cases h-index is less objective as judging the level of scholarship.

This paper aims at improving the flaws of h-index, mainly through establishing a new index based on its basic idea: the c-index, which is also supposed to be more suitable for the condition of China.

Principles of c‑index

In the academia, it is an important problem how to conveniently evaluate the level of scientific achievement of a scientist. The number of articles published can indicate something: The more the quantity, the higher the level of scientific achievement tends to be. Besides, the quality of an article is also significant, which is apparently often proportional to its number of citations. h-index is an index combining the considerations for both the quantity and quality of articles. However, h-index overemphasizes the lower number of citations. One knows that even a top scientist cannot assure that

each of his paper gains high citation rate. Actually, usually only several masterpieces are sufficient to set up the reputation of a top scientist.

Let us survey an example. Suppose that there are two scholars *A* and *B*, each with the same number of papers published. Suppose scholar *A* published 30 articles, with one quoted 100 times, and each of the rest is quoted 8 times. Meanwhile, the papers of scholar *B* are all quoted eight times, respectively. By computation, the h-index of *A* is identical to *B*, which is 8. Then it is taken for granted that the academic levels of two scholars are identical, according to h-index. However, from the point of view of common sense, this is not true, because one of the papers of *A* is extraordinarily excellent. Obviously this is unfair for *A*. To solve this problem, the h-index should be substantially improved.

Definition 1: Suppose an author has *n* papers published, each with the time of cites q_i $(i = 1, 2,..., n)$. The times of cites can be ordered as

$$q_1 > q_2 > ... > q_n \tag{1}$$

As a result, the c-index is

$$c = \min(\max_{(\sum_{i=1}^{k} q_i)/k \geq k} k, n) \tag{2}$$

This index is evidently more objective than h-index. In the above example, the c-index of scholar A is now 14, whereas the c-index of B is still 8. Thereby, the difference between the levels of these two scholars is clarified. The main idea of c-index is to replace the minimal times of cites for h-index by the average times of cites, as the gist of computation.

Roughly speaking, a simplified formula to compute c-index could be

$$c = \min\left(\sqrt{\sum_{i=1}^{n} q_i}, n\right) \tag{3}$$

Many scholars want to pursue high ranking of evaluation for their achievements. In order to improve academic reputation, some people tend to operate the h-index by trickery. A prevailing phenomenon is about the nominal authors. Typically several authors are listed in an article, but often the truth is that not everyone contributes to it. In fact, many research works are accomplished solo by the first author, with the others being just in the names. Many people thereby gain free authorship from their friends and thus have their h-indices improved. As an instance, a young author has 8 papers published, being totally cited 215 times, and his h-index displayed on Scopus is 6. However, actually he is the first author of only one of these papers. Thus, the h-index of him is overestimated, since the real h-index might be 1.

Because of the aforementioned reasons, many organizations in China only concern the works of a scholar as the first author and just neglect those co-authored works. Likewise, c-index should not be shared and should only count the first author. However, there still exists a common case that should be paid attention. Usually, a paper jointly authored by a student and his supervisor is actually written by the student, no matter whether he is the first author or not. But it is still always the truth that the supervisor should also have contributed a lot. In such a case, it would be unfair if the contribution of either the student or the supervisor is ignored. As a custom in China, there are two typical types of author sequences for a paper accomplished by student: (1) The student is nominated as the first author and the supervisor as the second and corresponding author; (2) The supervisor is nominated as the first author and the student as the second and corresponding author. It may worth mentioning that the current paper belongs to the second case. As a typical instance, Tsinghua University has officially released a rule, requiring the students to adopt the aforementioned first type of author sequence. A better measure for evaluating the academic level should not only ensure the benefit of the students, but also respect the guidance of the supervisors. Hence, the c-index ought to be improved comparing with h-index in this aspect. If the corresponding author is the second author, then we regard the relationship between the first two authors as student and supervisor. In this case, they should share the c-index

induced by the document, each with a specific percentage. Here we give a referential suggestion: The first author shares 2/3; the second and corresponding author shares the remaining 1/3. Otherwise, if the first author is the corresponding author, then he will get the full c-index points, whilst the other authors get zero.

In addition to the above problems, other issues remain through which the authors may get the h-index by trickery. For example, self-citation is a common phenomenon [10], which is apparently less significant than non-self-citations from other peers. Thus, it is reasonable to attach more importance to the non-self-citations. Nonetheless, non-self-citations could also be gained by trickery. It is possible that friends deliberately cite each other to increase quantity of non-self-citations of both sides for the sake of high h-index. As we know, such a behavior may be very effective to increase the h-index. Here we also give a referential suggestion: If a paper of an author is cited by his past cooperative partner, the cite counts 1/3, whereas pure self-citation counts nothing at all. Technically, this can be easily implemented through any intelligent author identification algorithm. Such rules could avoid the potential skillful schemes as much as possible.

Example

Let us take Prof. Jia, Qing-Shan with Tsinghua University as an example. His author profile is shown in Fig. 1, with an h-index displayed on scopus being 11. He has published 82 papers in total, and the total number of citations is 130. According to a thorough survey, it is a fact that he published altogether 19 articles as the first author, with 37 times of non-self-citations, sorted as follow

$$6, 6, 6, 6, 2, 2, 2, 2, 1, 1, 1, 1, 1$$

As the second and corresponding author, he has published seven articles, with 17 times of non-self-citations, sorted as

$$6, 3, 2, 2, 2, 1, 1$$

As a result, the c-index of him is almost 5.

Fig. 1 Author profile of Prof. Jia, Qing-Shan on Scopus

Let us take Prof. Cai, Ning with Northwest University for Nationalities as another example. His author profile is shown in Fig. 2, with h-index displayed on Scopus being 6. According to a survey in Scopus, he has published 23 papers in total, with the number of citations being 190. He published altogether 16 articles as the first author, with 49 times of non-self-citations, sorted as follow

$$18, 9, 9, 8, 5$$

As a result, the c-index of him is about 7, according to a rough calculation. One can see that, without counting the self-citations and non-first-author works, c-index would generally not be higher than h-index.

Fig. 2 Author profile of Prof. Cai, Ning on Scopus

Conclusion

Although the h-index enjoys many obvious advantages, and it has been widely accepted as a standard for evaluating the scientific achievement for individual scholar, it still bears many problems and needs to be substantially improved. In this paper, a new index is proposed, addressing to eliminate some drawbacks of h-index whilst inheriting the merits. This index is called c-index, which is named because it ought to be more suitable for the situation in China. The main idea of c-index is paying attention to the average time of cites. Besides, certain rational suggestions are also given for reference to further reinforce the objectiveness.

Acknowledgements

References

[1] Hirsch J. E., "An index to quantify an individual's scientific research output", *Proc. Natl. Acad. Sci.*, vol. 102, no. 46, pp. 16569-16572, 2005.

[2] Garcia-Perez M., "A multidimensional extension to Hirschs h-index", *Scientometrics*, vol. 81, no. 3, pp. 779-785, 2009.

[3] Bartneck C., Kokkelmans C., "Detecting h-index manipulation through self-citation analysis", *Scientometrics*, vol. 87, no. 1, pp. 85-98, 2011.

[4] "Impact of social sciences - 3: Key measures of academic influence", *Impact of Social Sciences*.

[5] Jones T., Huggett S., Kamalski J., "Finding a way through the scientific literature: Indexes and measures", *World Neurosurgery*, vol. 76, p. 36, 2011.

[6] Wendl M. "H-index: However ranked, citations need context", *Nature*, vol. 449, no. 7161, p. 403, 2007.

[7] Bornmann L., Daniel H. D., "What do citation counts measure? A review of studies on citing behavior", *Journal of Documentation*, vol. 64, no. 1, pp. 45-80, 2008.

[8]   Sekercioglu C. H., "Quantifying coauthor contributions", *Science*, vol. 322, no. 5900, p. 371, 2008.

[9] Zhang C.-T., "A proposal for calculating weighted citations based on author rank", *EMBO Reports*, vol. 10, no. 5, pp. 416-417, 2009.

[10] Aksnes D., "A macro study of self-citation", *Scientometrics*, vol. 56, no. 2, pp. 235-246, 2003.