

The Research on H-index-Based Academy Credit Evaluation Based on Fractional Theory

Hui-ying Gao*

School of Management, Hefei University of Technology, Hefei, 230039, China

Abstract: This paper studies the h-index-based academy credit evaluation based on fractional theory. The ACE model based on h-index was proposed and implemented. Learning from many study of academic performance evaluation, the paper proposed an ACE model emphasizing the credit from academic experts, with some new methods and indicators such as large-scale h-index extracting approach, citation recommendation analysis approach and z-index indicator. To apply the new model of ACE in real academic performance evaluation, related software was developed. The empirical study shows that the h-index-based academy credit evaluation based on fractional theory can achieve better performance than the traditional algorithm in academy credit evaluation.

Keywords: Academy credit evaluation, ACE model, h-index-based, real academic performance evaluation.

1. INTRODUCTION

Since the introduction of the h-index in 2005, the h-index has generally been well-received by the international scientific community and library and information science. The h-index research has been rapidly become one of frontiers in metrology information and academic evaluation. This paper reviewed the existing research results of h-index, then carried out empirical research to the topic of h-index of cross-data sources in scholars, academic journals and academic institutions three dimensions based on ISI web of science and Scopus two comprehensive abstracts citation database.

In the era of knowledge economy, academic performance evaluation has significant influence upon scholars, academic institutions, management institution and even countries. The result of academic performance evaluation is an important basis for research management and affects the development of science deeply [1, 2].

Therefore, such evaluation has important theoretical value and practical implications. There are many drawbacks in the current academic performance evaluation system, leading to serious crisis of trust. Moreover, the old academic performance evaluation framework based on citation analysis doesn't fit the changing publishing trends in the network environment [3, 4]. However, h-index could be different obtained from different data sources which will obviously increase the complexity and accuracy of academic assessment [5].

To solve the problems mentioned above, this paper is focused on the innovative treatment of basic concepts and methodology of academic performance evaluation, and uses h-index and related indicators as basic indicators for a new

academic performance evaluation model. This dissertation recommends a new model named Academic Credit Evaluation (ACE) so as to stress the academic qualification and credibility of the evaluators. It uses large-scale data extraction approach for h-index calculation and it has the advantage of measuring the career stage of every member in an academic community. Citation recommendation analysis approach is tried to explore the structure of an academic community, and z-index is suggested to measure the academic reputation of an author or a paper in high-impact groups.

The field of Information Science and The scientific evaluation of the academia have paid a lot of attention to H-index since it was brought up by professor Hirsch in 2005. Its theory has been improved and its application has also been broaden. However, this index has some deficiencies because it is calculated only by a single index called cited times. Bornmann's paper [6] starts from thesis some typical indices such as cited times, download times, author's h-index and periodical's in the year when thesis is published, annual average on cited times and download times using software SPSS to extract three principle components. The first contains cited times, download times, average on cited times and annual average on download times. The second contains authors' annual h-index. The third one contains periodical annual impact factor. Through analysing these three comprehensive principle component paper gets the six indices' weight which are 0.067-0.289, 0.157, 0.319, 0.291, 0.258. Paper gets the elected 20 authors' every thesis' total scores according to the six indices' weight and calculates 20 multiple h type indices Hm in accordance with the method which h-index is calculated by. The correlation analysis between Hm and h-index tells that these two indices' linear correlation is significant and the correlation coefficient reaches 0.771 which proves the Hm's feasibility. Meanwhile, in comparison with h-index, Hm which is researched on multiple indices has lower Equal probability, even if different

author had the same H_m we also can compare the authors by minimum and precise accurate decimal score. In addition, it's also useful to improve the unfair phenomenon between the young and old authors to some extent.

2. THEORETICAL DERIATLON

The standard conversion can reduce the difference between the numerical value of the periodical index, and it will be transformed into a simple numerical comparison. The h-index has received an enormous attention for being an indicator that measures the quality of researchers and organizations. The index value is compared with the relative position of the subject, so it can be avoided in the process of academic journal evaluation. The unfairness caused by the differences between the families, which make the journals of different disciplines more average, more fair, and the subject of the index more discipline. Then it can provide the basis for comparison. In this paper, through the analysis of the various indicators of the average value of the subject, regardless of the aim indicators, all disciplines of Science are discussed.

The average value of the family is very different, and the journals have different half-lives and citation density. Therefore, the direct comparison of these subjects is not fair; the standard in this paper is using the method of transformation. The average number can be recognized as the reference point, and the standard deviation is used as the unit to indicate the distance, and the relative position of the journals in the discipline is derived. Standard method is used for reflecting the relative influence of the discipline of the journal.

Academic evaluation is an important activity of management of academic output and is also a hot problem to which both scientists and evaluation institution pay much attention. In 2005, J. E. Hirsch, physics professor from the University of California, design of a new scientific evaluation index- h-index which improves the fairness and objective of academic evaluation. h-index was soon be well welcomed by all parties showing a good prospect of application. h-index is a new evaluation of academic indicators, immediately aroused wide interest and a high degree of concern in academic communities h-index as well as h-type index have become the core of possible next generation of evaluation parameters, so research on the h-index and the index is of great realistic significance and academic value.

Based on the significance of h-index, international and domestic electronic journal databases provide h-index-related inquiries and paper quoted data analysis functions, such as Web of Science citation databases and Scopus database. However, due to the difference of the recorded the types, quantity and time of journals, papers and references, h-index has discrepancies among different data sources which will increase the complexity of academic assessment [7].

Based on the research of h-index, we added the influence factor which combined with the fractional coefficient parameters to make some improvement on the h-index. No doubt H-index has improved the method to evaluate the scholar impact of a researcher, but it has some shortages. First, the performance of evaluating the impact and dominates of a scholar in a team when they have done a teamwork is not concerned in H-index. The H-index only concerns the

researches who take the main role in a research. For an example, a research noted as A research has been used as a part of a lot researches, but the author of A research is not the main researcher in all the others' research, if we use the H-index, the result can not affect the reality of the impact of research A.

Second, the h-index is not suitable for cross disciplinary comparison. Chinese readers are more familiar with the Hocking's h-index which is relatively high, is 62. Among them, the highest h-index is obtained by the Wolf prize for medicine, the professor of Johns Hopkins University, Schneider, the h-index is 161, followed by the Nobel Prize in physiology or medicine, California Institute of Technology, Baltimore, 160. The h-index is high, and it shows that the h-index is just like other indicators, not suitable for cross disciplinary comparison.

Third, the h-index contains the composition of self-citation components. This study recognized that influence factor of evaluation should be excluded from the self-citation components and prevent fraud case, making the results more effective and fair.

Based on the above points, we believe that we need to improve the H factor, especially for the individual contribution in team research, we need to distinguish and count the individual factors in order to make the evaluation more reasonable.

Recently in China in 2007, some academic evaluation institutions published the first h-index of scholarly journals, and a number of research institutions have put the h-index requirement into the hiring process as basic conditions. A person, a research institution or an academic journal has the higher h-index, the greater influence on the academic community. However, h-index could be different obtained from different data sources which will increase the complexity of academic assessment. Why a variety of data sources can cause the same author different h-index? Is there any method to reconcile the difference of h-index obtained from different data sources? In-depth analysis of the index and its formation mechanism in different data sources is the leading edge research and will be presented [8].

There have many drawbacks in the current academic performance evaluation system, leading to serious crisis of trust. Moreover, the old academic performance evaluation framework based on citation analysis doesn't fit the changing publishing trends in network environment [9, 10]. To solve the problems mentioned above, this paper is focused on the innovative treatment of basic concepts and methodology of academic performance evaluation, and use h-index and related indicators as basic indicators for a new academic performance evaluation model. This dissertation recommends a new model named Academic Credit Evaluation (ACE) so as to stress the academic qualification and credibility of the evaluators. It uses large-scale data extraction approach for h-index calculation and it has the advantage of measuring the career stage of every member in an academic community. Citation recommendation analysis approach is tried to explore the structure of an academic community, and h-index is suggested to measure the academic reputation of an author or a paper in high-impact groups.

Many scholars proposes an organized comparison of several academic research groups under the same disciplines, according to Hirsch (*h*) based the bibliometric indicators. For the comparison to be as much full as possible, several different typologies of *h*-based indicators are taken into use. They are respectively: the *h*-spectrum, the *h*-index of a research group (h_{GROUP}), the successive *h*-index of a research group (h_2) and the *h*-index of single publications (h_{SINGLE}).

Li has presented a methodology for comparing homologous research groups on the basis of their scientific production and academic credit. Through the above analysis of correlation, *h*-index, *g*-index are both close with the traditional citation evaluation. The correlation coefficient is very big. And the *h*-index in evaluation of colleges and universities and scholars at some discipline is well within the influence of comprehensive scientific output quality and quantity. But the *g*-index has inherent advantages in *h*-index on the basis of further optimize the college academic performance evaluation index, to make it more scientific and reasonable. Assessment of academic performance evaluation is a key index of the academic level of scientific research workers, which is also one of the important indicators to measure the level of higher education. Therefore, that adopting the scientific academic evaluation index to evaluate academic university scholar, has the extremely important practical significance [11].

Firstly, the problem data update is the problem. The subject line must be regularly updated. Secondly comes the data identifying problem. In the process of evaluation, large amount of data will be dealt. The accuracy of the data should be ensured in the process. Thirdly, it is the database selection problem. Evaluation data is based on all kinds of database. Therefore, in the actual evaluation should be based on subject classification to selection, database to make evaluation result more scientific and reasonable.

3. THE FRAMEWORK OF FRACTIONAL THEORY

It is necessary for the *h*-index that the number of citations per article is not smaller than *h*. However, it is not necessary for the *h*-index that the number of citations per article is not smaller than *g*.

The *h*-index has been characterized in terms of three natural axioms by Woeginger. The simplest of these three axioms states that by moving citations from weaker articles to stronger articles, one's research-index should not decrease. Assume that a researcher has published *m* articles with a sequence of citation numbers ranked in decreasing order:

$$n_1 \geq n_2 \geq n_3 \geq \dots \geq n_k \geq \dots \geq n_m \geq 0$$

where $n_j (j = 1, 2, \dots, m)$ are the cited times of the *j*-th publication respectively. Then the *h*-index is the unique largest publication number such that the top *g* articles received together the cumulative citations $n_1 + n_2 + \dots + n_g \geq g^2$. In other words, we have:

$$\frac{n_1 + n_2 + \dots + n_g}{g} \geq g \tag{1}$$

Geometrically, the *h*-index means the largest width of all the rectangles with the bottom (width) of publications and the height (length) of the cumulative citations with at least the square of the bottom (width) of publications.

The *h*-index is a new index in citation analysis for measuring and comparing the output of a scientific researcher, which may be considered as a variation of the *h*-index.

Assume that a researcher has published *m* articles with a sequence of citation numbers ranked in decreasing order:

$$n_1 \geq n_2 \geq n_3 \geq \dots \geq n_k \geq \dots \geq n_m \geq 0$$

where $n_j (j = 1, 2, \dots, m)$ represent the *j*-th paper has been cited by others at least n_j times, respectively.

Definition 3: Given a set of articles by a researcher ranked in decreasing order of the number of citations that they received, the *f*-index is the largest one among all the geometric averages of each publication and the corresponding citation. Mathematically, *f*-index can be expressed by

$$f = \max_{1 \leq j \leq m} \sqrt{jn_j} \tag{2}$$

According to the above mentioned definition, we can easily calculate the *f*-index from Table 1.

Table 1. The H-index calculation.

Rank	Citation	Geometric Average	<i>f</i> -index
1	n_1	$\sqrt{1n_1}$	
2	n_2	$\sqrt{2n_2}$	
...	
$k - 1$	n_{k-1}	$\sqrt{(k - 1)n_{k-1}}$	
k	n_k	$\sqrt{kn_k}$	$f = \max_{1 \leq j \leq m} \sqrt{jn_j}$
$k + 1$	n_{k+1}	$\sqrt{(k + 1)n_{k+1}}$	
...	
m	n_m	$\sqrt{mn_m}$	

Assume that a researcher has published *m* articles with a sequence of citation numbers ranked in decreasing order:

$$n_1 \geq n_2 \geq n_3 \geq \dots \geq n_k \geq \dots \geq n_m \geq 0$$

where $n_j (j = 1, 2, \dots, m)$ are the cited times of the *j*-th publication, respectively. Geometrically, the *f*-index is the arithmetic square root of the area of the largest inscribed rectangle among all the rectangles with the bottom (width) of publications and the height (length) of citations.

Each citation index of those indices has its own strengths and weaknesses. The *h*-index has several good properties as follows:

1. The *f*-index is a single indicator which takes into accounts both the quantity and impact of the publications.
2. It is simple to compute.

Table 2. The productive researcher with the list of publications and citations.

Rank	Citation	Geometric Average	f-index
1	n_1	$\sqrt{n_1}$	$\bar{f} = \max_{1 \leq j \leq m} \frac{\sqrt{n_1} + \sqrt{2n_2} + \dots + \sqrt{jn_j}}{j}$
2	n_2	$\frac{\sqrt{n_1} + \sqrt{2n_2}}{2}$	
...	
k	n_k	$\frac{\sqrt{n_1} + \sqrt{2n_2} + \dots + \sqrt{kn_k}}{k}$	
...	
m	n_m	$\frac{\sqrt{n_1} + \sqrt{2n_2} + \dots + \sqrt{mn_m}}{m}$	

Table 3. The list of publications and citations.

Publication Rank	Citation Sequence	Cumulative Citation	Geometric Average	Arithmetic Average
1	13	13	$\sqrt{13} = 3.61$	3.61
2	9	22	$\sqrt{18} = 4.24$	3.93
3	6	28	$\sqrt{18} = 4.24$	4.03
4	4	32	$\sqrt{16} = 4.00$	4.02
5	3	35	$\sqrt{15} = 3.87$	3.99
6	0	35	$\sqrt{0} = 0$	3.33
indices	$h = 4$	$g = 5$	$f = 4.24$	$\bar{f} = 4.03$

- It takes into account the cites of the highly cited papers (but the h-index is insensitive to highly cited papers).
- It provides more granularity than the h-index. As we see popularly, to increase the h-index is difficult (especially when the h-index is high) and it is usual to find that many different researchers have the same h-index with a very different number of publications and cites.

However, the h-index has its own drawbacks. It is not easy to understand the f-index since it is valued in the scale of geometric mean. Moreover, it is not robust in some cases.

For the indicator which became better and bigger:

$$z(i, j) = S_{\min} + \frac{e(i, j) - \min_j e(i, j)}{\max_j e(i, j) - \min_j e(i, j)} \times (S_{\max} - S_{\min}) \tag{3}$$

For the indicator which became better and smaller:

$$z(i, j) = S_{\max} - \frac{e(i, j) - \min_j e(i, j)}{\max_j e(i, j) - \min_j e(i, j)} \times (S_{\max} - S_{\min}) \tag{4}$$

For e-plus in some appropriate value indicators:

$$z(i, j) = S_{\max} - \frac{|e(i, j) - e_0|}{|e(i, j) - e_0| + e_0} \times (S_{\max} - S_{\min}) \tag{5}$$

In the evaluation of the ICP, it can be estimated the edge nodes for selecting is that:

$$(cm(k, j))_{n_c \times m_a} = \frac{\sum_{i \in pc(k, j)} w_i}{\sum_{i=1}^{n_c} w_i} \tag{6}$$

$$pc(k, j) = \{c_i | (a_k \succ a_j) \cup (a_k \sim a_j)\} \tag{7}$$

4. EXPERIMENT RESULTS

In the following we present a few examples for use of the h-index in the evaluation of the scientific output of researchers.

Test 1: Consider the examples in Table 2 where number of publications and citations are shown. Results show that the h-index, h-index, and f-index may be different. This example shows the good behaviour of h-index.

Test 2: Suppose that we want to compare the scientific production of two different researchers. The first one has published 15 papers and the second researcher has also published 15 papers. According to the definition of h-index, both have an h-index of 11. But their 11 most cited papers have received different sites for each paper. It is obvious that the production of the second researcher has a higher impact factor. The h-index of the second researcher is higher than that of the first one. This is not reasonable. The f-index of the second researcher is lower than that of the first one.

Table 4. The productive researcher with the list of publications and citations.

Paper Rank	Citation Number	Rectangle Area	Geometric Average	<i>f</i> -index	\bar{f} -index
1	59	59	7.68		
2	50	100	10.00		
3	30	90	9.49		
4	24	96	9.80		
5	20	100	10.00		
6	20	120	10.95		
7	17	119	10.91		
8	17	136	11.66	$f_A = 11.66$	
9	14	126	11.22		
10	12	120	10.95		
11	11	121	11.00		
12	9	108	10.39		$\bar{f}_A = 10.34$
13	8	104	10.20		
14	6	84	9.17		
15	6	90	8.94		

Table 5 Publications and citations.

Rank	Citation	Rank	Citation	Rank	Citation	Rank	Citation
1	909	26	79	51	35	76	14
2	897	27	77	52	35	77	12
3	799	28	76	53	34	78	12
4	306	29	76	54	33	79	12
5	283	30	75	55	33	80	11
6	276	31	71	56	29	81	10
7	251	32	71	57	27	82	10
8	223	33	71	58	26	83	9
9	219	34	65	59	25	84	8
10	210	35	64	60	24	85	8
11	186	36	54	61	24	86	6
12	165	37	52	62	23	87	6
13	162	38	52	63	23	88	6
14	141	39	51	64	22	89	5
15	128	40	50	65	21	90	5
16	114	41	50	66	21	91	5
17	110	42	49	67	20	92	5
18	109	43	48	68	18	93	3
19	105	44	46	69	18	94	2
20	105	45	46	70	18	95	2
21	103	46	41	71	17	96	1
22	98	47	41	72	17	97	1
23	94	48	39	73	14	98	0
24	93	49	39	74	14	99	0
25	85	50	36	75	14	100	0

Test 3: Examine a productive researcher with the list of publications and citations shown in Table 5 based on Google Scholar. Then the h-index is 45, the h-index is 91, the h-index is 48.96, and the h-index is 44.52.

Fig. (1) shows the geometrical interpretation of the h-index and Table 3 shows the list of publications and citations. Table 4 shows the h-index and h'-index Calculations of Author A.

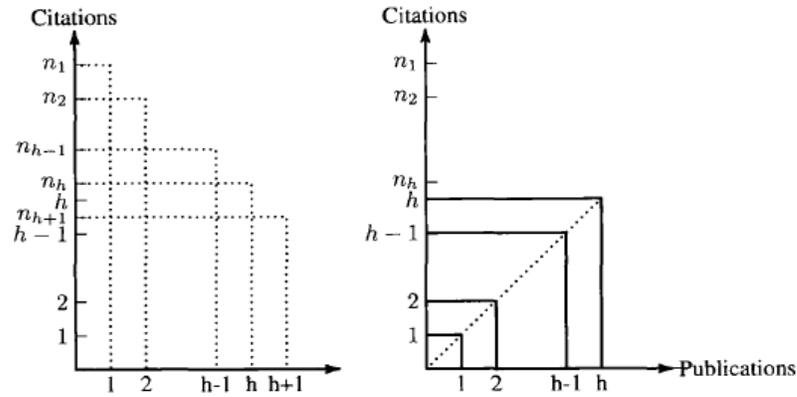


Fig. (1). The geometrical interpretation of the h-index.

CONCLUSION

In the last years, the h-index, a measure of the scientific output of researchers based on both the quantity and impact of publications, has received great attention from the scientific community. Many papers have dealt with this index and have proposed new variations of the h-index (for example, the h-index).

Regarding the future, our analysis will be extended to other scientific sectors at national or even international level. Also, we will evaluate the opportunity of aggregating results related to different scientific sectors into a general ranking, able to reflect the overall scientific production of universities or research institutions.

CONFLICT OF INTEREST

The author confirms that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

- [1] R. Arencibia-Jorge, I. Barrios-Almaguer, S. Fernandez-Hernandez and R. Carvajal-Espino, "Applying successive h indices in the institutional evaluation: A case study" *Journal of the American Society*

for Information Science and Technology, vol. 59, no. 1, pp. 155-157, 2008.

- [2] M.G. Banks, "An extension of the Hirsch-index: Indexing scientific topics and compounds", *Scientometrics*, vol. 69, no. 1, pp. 161-168, 2006.
- [3] J. Bar-Ilan, "Which h-index? A comparison of WoS, Scopus and Google Scholar", *Scientometrics*, vol. 74, no. 2, pp. 257-271, 2008.
- [4] P.D. Batista, M.G. Campiteli, O. Kinouchi, and A.S. Martinez, "Is it possible to compare researchers with different scientific interests?", *Scientometrics*, vol. 68, no. 1, pp. 179-189, 2006.
- [5] C. Bergstrom, "igenfactor: Measuring the value and prestige of scholarly journals," *College and Research Libraries News*, vol. 68, no. 5, pp. 314-316, 2007.
- [6] L. Bornmann, R. Mutz, and H.D. Daniel, "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," *JASIST*, vol. 59, no. 5, pp. 830-837, 2008.
- [7] G. Zhou, A. Cichocki, Q. Zhao, and S. Xie "Nonnegative matrix and tensor factorizations: an algorithmic perspective," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 54-65, 2014.
- [8] Q. Burrell, "Hirsch-index or Hirsch rate? Some thoughts arising from Liang's data," *Scientometrics*, vol. 73, pp. 19-28, 2007.
- [9] D. Butler, "Free journal-ranking tool enters citation market," *Nature*, vol. 451, no. 7174, pp. 6, 2008.
- [10] H. Ceylan, and H. Ceylan, "A hybrid harmony search and TRANSYT hill climbing algorithm for signalized stochastic equilibrium transportation networks," *Transportation Research Part C: Emerging Technologies*, vol. 25, pp. 152-167, 2013.
- [11] R. Arul, G. Ravi, and S. Velusami, "Chaotic self-adaptive differential harmony search algorithm based dynamic economic dispatch," *International Journal of Electric Power Energy System*, vol. 50, pp. 85-96, 2013.

Received: June 10, 2015

Revised: July 29, 2015

Accepted: August 15, 2015

© Hui-ying Gao; Licensee Bentham Open.

This is an open access article licensed under the terms of the (<https://creativecommons.org/licenses/by/4.0/legalcode>), which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.