Jointly published by Akadémiai Kiadó, Budapest and Springer, Dordrecht

Scientometrics, Vol. 69, No. 1 (2006) 153–159

h-index sequence and *h*-index matrix: Constructions and applications

LIMING LIANG[a,b]

[a] *Institute for Science Technology and Society, Henan Normal University, Xinxiang (P. R. China)*
[b] *University of Antwerp (UA), IBW, Wilrijk (Belgium)*

The calculation of Hirsch's *h*-index is a detail-ignoring way, therefore, single *h*-index could not reflect the difference of time spans for scientists to accumulate their papers and citations. In this study the *h*-index sequence and the *h*-index matrix are constructed, which complement the absent details of single *h*-index, reveal different increasing manner and the increasing mechanism of the *h*-index, and make the scientists at different scientific age comparable.

Introduction

Months ago J. E. Hirsch proposed the index *h*, defined as the number of papers with at least *h* citations each, as an index to measure the scientific output of a researcher (HIRSCH, 2005). A novel and interesting indicator, *h*-index has been discussed or developed by some studies. A short paper published in *Nature* made the *h*-index known to many scientists (BALL, 2005). Braun and his colleagues used the *h*-index in the citation assessment of journals (BRAUN et al., 2005). Van Raan presents characteristics of the statistical correlation between the *h*-index and several standard bibliometric indicators, as well as a comparison with the results of peer review judgment

Received March 7, 2006

Address for correspondence:
LIMING LIANG
Institute for Science Technology and Society, Henan Normal University
Xinxiang, 453007 P. R. China
E-mail: pllm@public.xxptt.ha.cn

0138–9130/US $ 20.00

(VAN RAAN, 2005). The *h*-index is also applied to distinguish between successful and non-successful applicants for post-doctoral research fellowships (BORNMANN et al., 2005).

Single *h*-index restricts itself within a static and uniform status. When Hirsch calculated the *h*-indexes of the 19 physicists, he only put the total numbers of the papers and citations of each scientist under considerations, ignoring the different time spans for those scientists to accumulate their papers and citations. Suppose two scientists have the same *h*-index, while one's academic career is much shorter than that of the other, what does it mean? Single *h*-index could not reflect such difference and reveal something behind this phenomenon. Therefore, Hirsch's index would become more active and useful if we could find a way to show the calculation background (or counting conditions) of a group of *h*-indexes and the variation of the *h*-indexes along with the changing of the calculation background. To do so, we propose a tentative method by constructing the sequence of the *h*-indexes (here after *h*-sequence for short) and the matrix of the *h*-indexes (*h*-matrix for short).

This method is composed of two steps:

1. Calculate the *h*-sequence by continually changing the time spans of the data;
2. Construct the *h*-matrix based on a group of correlative *h*-sequences.

The papers and citation records of the 19 physicists mentioned in Hirsch's case (HIRSCH, 2005) were searched at ISI's Web of Science on 27 Sep. 2005 with a time-span from 1955 to 2004. Based on the records and Hirsch's method we computed 19 *h*-indexes of the physicists. Among them, only eleven are accorded with the *h*-indexes calculated by Hirsch, largely due to that the publication and citation data of the physicists as well as the records in the database had changed during the interval of the two researches. In order to collate our *h*-sequences with Hirsch's *h*-index, we just selected these eleven physicists as our sample set to illustrate how to create the *h*-index sequence and matrix and what implications they have.

How to construct the *h*-index sequence and *h*-index matrix

In Hirsch's paper E. Witten is the physicist with the highest *h*-index 110. We take Witten's data as an example to explain how to create an *h*-index sequence. The records show that Witten's first paper was published in 1976. According to Hirsch's definition of the *h*-index, based on the number of Witten's papers published in year 2004 and the number of citations earned after the papers' publication, we calculate Witten's *h*-index for year 2004, denoted as h_1. Here, h_1=3. Based on the number of papers published in 2003 and 2004 and the number of their citations received in 2003 and 2004, we obtain Witten's *h*-index for the period 2003–2004, denoted as h_2. Here, h_2=7. Similarly, we calculate Witten's *h*-indexes for the period 2002–2004, 2001–2004, … , 1976–2004, and denoted as h_3, h_4, …, h_{29}. By this way we create Witten's *h*-sequence, including 29 h_i-indexes (*i* is the number of the covered publication and citation years).

By the same way, we create all *h*-sequences of the eleven physicists, then we arrange all the 11 *h*-sequences in a matrix as shown in Table 1 and we call the matrix "the *h*-matrix". In Table 1, there are nine physicists with the *h*-sequences consisting of more h_i-indexes than Witten's *h*-sequence. There is a bolded *h*-indexes in each *h*-sequence in the matrix, which is nothing but the *h*-index shown in Hirsch's original paper and is denoted as *H*-index in this paper.

In Table 1, we do not use the real names of the physicists, just denote as physicist 1, physicist 2, and so on. It is because we have not identified whether in the searched records there are the records belonging to other scientists who have the same names as some of the eleven physicists. In addition, our intention here is not to evaluate the physicists' scientific achievements by using *h*-index, but to select samples to explain the construction and application of the *h*-sequence and *h*-matrix.

Figure 1 presents the 11 *h*-sequences.

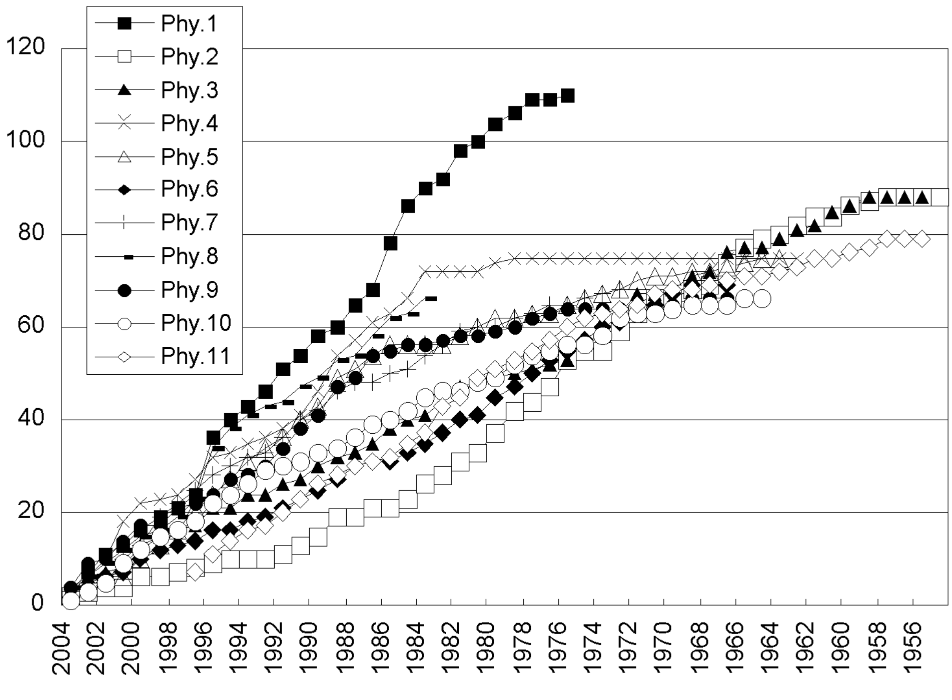

Figure 1. *h*-sequences of the 11 physicists

Table 1. *h*-index sequences and *h*-index matrix of eleven physicists

	i	Phy.1	Phy.2	Phy.3	Phy.4	Phy.5	Phy.6	Phy.7	Phy.8	Phy.9	Phy.10	Phy.11
2004	1	3	2	3	3	3	1	3	4	4	1	
2003	2	7	3	6	6	5	4	4	6	9	3	
2002	3	11	4	7	10	6	6	5	9	10	5	
2001	4	13	4	8	18	6	7	10	12	14	9	
2000	5	16	6	13	22	10	10	13	15	17	12	
1999	6	19	6	15	23	13	12	19	16	17	15	
1998	7	21	7	16	24	16	13	22	19	21	16	
1997	8	24	8	17	27	21	14	25	23	22	18	7
1996	9	36	9	21	32	24	16	28	34	24	22	11
1995	10	40	10	21	33	27	16	30	38	27	24	14
1994	11	43	10	24	35	31	18	32	41	28	26	16
1993	12	46	10	24	36	34	19	33	43	30	29	17
1992	13	51	11	26	38	36	21	36	44	34	30	20
1991	14	54	13	27	41	40	23	40	47	38	31	23
1990	15	58	15	30	46	43	25	42	49	41	33	26
1989	16	60	19	32	54	49	27	46	53	47	34	28
1988	17	65	19	33	57	51	30	48	54	49	36	30
1987	18	68	21	35	61	54	31	48	58	54	39	31
1986	19	78	21	38	63	56	31	50	62	55	40	32
1985	20	86	23	40	66	56	33	51	63	56	42	35
1984	21	90	26	41	72	56	35	54	**66**	56	45	37
1983	22	92	28	44	72	56	37	56		57	46	43
1982	23	98	31	47	72	58	40	59		58	46	45
1981	24	100	33	48	72	60	41	60		58	48	49
1980	25	104	37	49	74	62	45	60		59	49	51
1979	26	106	42	50	75	62	47	62		60	52	53
1978	27	109	44	51	75	63	50	63		62	54	55
1977	28	109	47	52	75	63	53	65		63	55	57
1976	29	**110**	53	53	75	65	55	65		64	56	60
1975	30		55	56	75	66	57	66		64	56	62
1974	31		55	59	75	67	60	67		64	58	63
1973	32		59	63	75	68	61	**68**		64	62	64
1972	33		63	67	75	70	63			64	63	65
1971	34		66	67	75	71	65			65	63	67
1970	35		66	68	75	71	67			65	64	68
1969	36		68	71	75	72	69			66	65	68
1968	37		69	72	75	72	69			66	65	69
1967	38		74	76	75	73	**69**			**66**	65	70
1966	39		77	77	75	74					66	71
1965	40		79	77	75	75					**66**	71
1964	41		80	79	75	**75**						72
1963	42		82	81	**75**							73
1962	43		84	82								75
1961	44		84	85								75
1960	45		86	86								76
1959	46		87	88								77
1958	47		88	88								79
1957	48		88	88								79
1956	49		88	**88**								**79**
1955	50		**88**									

Application of the *h*-index sequence and *h*-index matrix

From the analysis of each *h*-sequence and the comparison of the *h*-sequences in the *h*-matrix we obtain certain useful information.

First, *h*-sequence reveals different increasing manner and the increasing mechanism of the *h*-index. In Hirsch's paper he wrote: "For a given individual one expects that *h* should increase approximately linearly with time." However, this is an ideal increasing type, and the simplest possible model is "assume the researcher publishes *p* papers per year and each published paper earns *c* new citations per year every subsequent year". The practical productivity is not so regular. Observing the curves in Figure 1 we found that the *h*-sequences have different increasing manner. The *h*-sequence of Phy.1 is indeed increasing linearly; The *h*-curve of Phy.2 is more like an "s" curve; The *h*-curve of Phy.4 seems like the Lorenz curve with a ceiling. While the left beginning year of the curve of Phy.11 is 1997, not 2004 as other ten curves, showing that after 1997 Phy.11 stopped publishing. These different increasing types provide clues to dig into the changing mechanism of the *h*-index. Generally speaking, during the rapid increasing period of the *h*-index, the individual's scientific production as well as his/her academic impact, is also active. In this paper we will not discuss the changing mechanism in detail.

Second, the *h*-matrix makes the scientists at different scientific age comparable. In the *h*-index matrix shown in Table 1 we found that Phy.2 and Phy.3 have the same *H*-index 88, Phy.4 and Phy.5 are all with H=75, and Phy.8, Phy.9 and Phy.10 share the *H*-index 66. If taking the year publishing the first paper as the beginning of a scientist's academic career, obviously, the scientific ages of the 11 physicists are not the same. In this case, when we only compare their *H*-indexes, it is difficult to make judgment of whose achievements are better. However, the *H*-indexes could be comparable in the *h*-matrix by taking a certain year as the beginning year of the *h*-sequences of all the scientists who published the first paper no late than this year. For example when the year is set as 1976 (labeled by the shadows in the *h*-matrix), in this case the *h*-index of Phy.2 is 26, smaller than the *h*-indexes of all other physicists, though the *H*-index of Phy.2 is as high as 88, ranking the second of the 11 physicists. In view that this cutting method could be "unfair" to elder scientists (such as Phy.2, Phy.3, Phy.11) as the period after1976 is not their high productive period, another measure could be adopted. We could choose the first *n* years in every scientist's academic career to calculate the *h*-sequence and then construct a new *h*-matrix. Based on the new *h*-matrix we could make a new comparison. However, the problem may still exist. For example, given that the first 30 academic years of Phy.2 is 1955–1984 while that of Phy.1 is 1976–2005, the comparison is still not totally fair, as the publication and citation situation keeps changing over time. Nevertheless, *h*-sequence and *h*-matrix identify these issues and ask us to find the way to solve these problems.

Conclusion and discussion

In determining the *h*-index of a scientist one just focuses on a section of the citation ranking list of the scientist's articles, i.e. the section near rank *h*. One does not mind about the complete ranking. Therefore, the calculation of the *h*-index is a detail-ignoring way. The *h*-sequence and *h*-matrix could complement the absent details, reveal different increasing manner and the increasing mechanism of the *h*-index, and make the scientists at different scientific age comparable. At the same time Hirsch's original *h* index could also find its position in *h*-sequence and *h*-matrix, i.e. the *H*-index.

h-sequence and *h*-matrix offer us some clues to consider how to use *h*-index more reasonably. One of our considerations is using the first *n* years of a scientist's academic career as the time-span of the calculation of the *h*-sequence. Here, *n* may equal to 10, 15, 20, and so on, The beginning year could be the year when publishing the first paper, or the year when receiving his/her PhD. This will be one of our future studies. Another attempt will be to select the most productive *n* years, or the most active *n* years of a scientist as the time-span to calculate the *h*-sequence, then to compare the *h*-sequences of the scientists at different academic ages. Related to these two designs, however, another problem emerged: how to restrict the citation window? In general, the earlier the paper published, the longer the citable period would be. So, when we determine the *n* years as the examined period, the citation window of the papers published during this period should be normalized as well. For example, taking *m* as the length of the citation window. For all the papers we just count their citations received since the publication year until the m^{th} year after its publication.

A more difficult problem is, when we use *h*-index as an indicator to measure the research performance of researchers, how can we eliminate the influence of database size, or we say the number of the records of the database, on the measure? We know, if a database contains more source journals, its records would also increase. Taking SCI as an example, the number of documents covered by SCI has been increasing linearly over the past fifty years (LIANG et al., 2005). Therefore, for a paper published in 1990 it is possible for us to search more citations from SCI 1995 than the citations searched from SCI 1985 received by a paper published in 1980, though both 1985 and 1995 are the 6^{th} year after the paper's publication. The solution of this problem is in consideration.

*

The author thanks the reviewer for the carefully reading and the useful suggestions. The work presented in this paper was supported by the National Natural Science Foundation of China (grant no. 70373055).

References

BALL, P. (2005), Index aims for fair ranking of scientists. *Nature,* 436 : 900.

BORNMANN, L., H.-D. DANIEL (2005), Does the h-index for ranking of scientists really work? *Scientometrics,* 65 (3) : 391–392.

BRAUN, T., W. GLÄNZEL, A. SCHUBERT (2005), A Hirsch-type index for journals. *The Scientist*, 19 (22) : 8.

HIRSCH, J. E. (2005), An index to quantify an individual's scientific output. *Proceedings of the National Academy of Sciences of the United States of America*, 102 : 16569–16572.

LIANG, L., R. ROUSSEAU, F. SHI (2005), The Rhythm of science, the rhythm of *Science*. *Proceedings of the 10th International Conference of the International Society for Scientometrics and Informetrics, Stockholm*, published by Karolinska University Press,

VAN RAAN, A. F. J. (2005),Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups, *Scientometrics*, 67 (1) 491–502.