New developments related to the Hirsch index

Ronald Rousseau

KHBO (Association K.U.Leuven), Industrial Sciences and Technology, Zeedijk 101, 8400 Oostende, Belgium

The Hirsch index: definitions

The Hirsch index (in short h-index) was introduced by Jorge E. Hirsch (2005) as an indicator for lifetime achievement, as measured by the number of received citations. More precisely, Hirsch defined the h-index as follows: *A scientist has index h if h of his/her papers have at least h citations each, and the other papers have fewer than h citations each.* It was soon pointed out (Glänzel, 2006) that this definition is not quite precise. Indeed, consider Table 1a.

Table 1a. Short fictitious table for determining the h-index

Publication	Number of citations
A	5
B	3
C	3
D	3
E	1

According to Hirsch' original definition the h-index for Table 1 is not 4 because it is not true that there are 4 articles with at least 4 citations. The h-index is not 3 either because it is not true that there are 3 articles with at least 3 citations (this part is logically correct) and the other ones have fewer than 3 citations (this part is now false). Consequently a precise definition of the h-index goes as follows.

Consider the list of publications [co-]authored by scientist S, ranked according to the number of citations each of these has received. Publications with the same number of citations are given different rankings (the exact order does not matter, anti-chronologically might be a good choice). Then S' Hirsch index is h if the first h publications received each at least h citations, while the publication ranked h+1 received strictly less than h+1 citations. Stated otherwise: S' Hirsch index is h if h is the highest rank (largest natural number) such that the first h publications received each at least h citations. Applying this definition to Table 1b yields unambiguously that the h-index is equal to 3.

Table 1b. Short fictitious table for determining the h-index, with ranking added

Publication (date)	Rank	Number of citations
A (2001)	1	5
B (2005)	2	3
C (2004)	3	3
D (2000)	4	3
E (2006)	5	1

The reason for suggesting anti-chronological ranking is that I would like to propose to call *Hirsch core* the set consisting of the first h articles, see also (Egghe et al.,2006). When using this ranking younger articles have a large probability to belong to the Hirsch core than older ones. It is further noted that if the last article in the list occupies rank R and receives C > R citations then this scientist's h-index is equal to R.

Although the h-index is a relatively simple indicator it attracted a lot of attention (Ball, 2005; Bar-Ilan, 2006; Batista et al.,2005; Cronin & Meho, 2006; Egghe, 2006c; Glänzel & Persson, 2005; Liang, 2006). It was very fortunate that the new journal *Science Focus* was able to communicate this excitement through a special section devoted to the h-index. Colleagues Glänzel, Jin, Moed, Van Raan and myself discussed advantages and disadvantages of this new indicator. It became also clear that the h-index can not only be used for lifetime achievements, but also in the context of many other source-item relationships (Braun et al., 2005; Egghe & Rousseau, 2006). In my contribution to *Science Focus* I noted, however, that not all source-item relationships fit into the calculation scheme for the h-index, by giving the example of bowling scores. A better example is the case of cities as sources and their population as 'produced items'.

A slight generalization of the h-index as defined above is obtained as follows. Let us denote the number of citations received by the article ranked r (in general source-item terminology: the source ranked r), as P(r), and its piecewise linear interpolation as P(x), this is: the function connecting the points (r, P(r)), where r denotes the rank (r = 1, 2, ...). Then the h-index may be defined as the abscissa of the intersection of the lines y = x and the observed function P(x). The original h-index is always a strictly positive integer, while this generalization, denoted as h_r, is a real number. Note that h_r is an index derived from observed data. If h_r is known than the corresponding h-value is equal to $\lfloor h_r \rfloor$. This is the floor function of h_r, or the largest natural number smaller than or equal to h_r. I suggest that using a real-valued Hirsch index is the natural thing to do when, e.g., citations are counted fractionally (Rousseau, 2006).

Another look at the same ranking: the g-index and the A-index.

It has been observed that (in the original context of publications and citations) the h-index is only weakly sensitive to the number of citations received. Indeed, when a scientist's h-index is equal to h then this scientist's first h articles received at least h times h, i.e. h^2 citations. This lower bound is the only relation that logically exists between publications and citations, when the h-index is known. There is no upper bound for the number of citations received by the Hirsch core, this is the set of articles occupying the first h ranks.

For this reason Leo Egghe proposed another index, referred to as the g-index (Egghe, 2006a,b). Also Jin Bihui, main editor of *Science Focus* proposed a way to take the number of citations into account (Jin, 2006). Her proposal is to use the average number of citations of articles in the Hirsch core as a more sensitive indicator. As this proposal uses an average I will refer to it as the A-index. Clearly, the A-index is usually a positive real number (not necessarily an integer).

The g-index is calculated as follows: one draws the same list as for the h-index, but now the g-index is defined as the highest rank such that the cumulative sum of the number of citations received is larger than or equal to the square of this rank. Clearly $h \leq g$ and $h \leq A$.

The g-index too can be calculated as a real number. It is then defined as the abscissa of the intersection of the curves $y = x^2$ and $y = C(x)$, where C(x) is the function connecting the points $C(r) = \sum_{k=1}^{r} P(k)$. Similar to the notation h_r, this index is denoted as g_r.

Consider now Table 2. S1's h-index is equal to 6, his g-index is equal to 8 because $64 = 8^2 \leq 70$, while $81 = 9^2 > 75$. The corresponding value of the A-index is 59/6 = 9.83.

Table 2: Publication and citation list of scientist S1

Rank (squared) - Publications	Citations	Sum
1 (1) A	20	20
2 (4) B	10	30
3 (9) C	9	39
4 (16) D	8	47
5 (25) E	6	53
6 (36) F	6	59
7 (49) G	6	65
8 (64) H	5	70
9 (81) I	5	75

Considerations regarding these three related indices

The h-index does not take articles receiving a small number of citations into account. This is a well-known advantage of this measure. It, moreover, mainly reflects the number of highly cited articles, but the actual number of citations does not influence the value of the h-index. A scientist who writes many articles which are each well-received, but not exceptionally well, will have a high h-index. His g-index will just be marginally larger than his h-index. Stated otherwise the ratio of g/h will be close to 1 (but never smaller than 1!). A scientist who writes a few exceptional articles, while her other articles are hardly noticed by the scientific community will have a relatively low h-index and a high g-index. Table 3 gives data for such extreme cases.

Table 3: More examples: publication and citation lists of scientists S2, S3, S4 and S5.

Rank (squared) - Publications	Citations (S2 – S3 – S4 –S5)	Partial sum (S2-S3-S4-S5)
1 (1) A	10 30 6 80	10 30 6 80
2 (4) B	9 20 6 5	19 50 12 85
3 (9) C	8 15 6 5	27 65 18 90
4 (16) D	7 9 6 2	34 74 24 92
5 (25) E	6 7 6 2	40 81 30 94
6 (36) F	6 6 6 2	46 87 36 96
7 (49) G	6 4 6 1	52 91 42 97
8 (64) H	5 2 6 1	57 93 48 98
9 (81) I	5 1 6 1	62 94 54 99
10 (100) J	4 1 6 1	66 95 60 100

Scientists S2, S3 and S4 have the same h-index as S1 (namely 6), yet their A- and g-indices differ: Their A-indices are 7.67, 14.5 and 6, while their g-indices are: 7, 9 and 6. Recall that S1 has a g-index equal to 8 and an A-index equal to 9.83. These scientists have either less (S2 and S4)) or more (S3) citations than the first one. Such differences are reflected by the A- and the g-index. S5 is a special case: this scientist has one highly-cited article, but his other articles did not receive a lot of citations. His h-index is only 3, while his g-index is still 10; his A-index is 30.

Conclusion

These examples show that the h-index on the one hand, and the A- and g-indices on the other, measure different things. The A-index seems overly sensitive to one extremely highly cited article (the case of scientist S5). For this reason it would seem that the g-index is the more useful of the two. As to the h- and the g-index: they do measure different aspects of a scientist's publication list. Certainly the h-index does not tell the full story, and, although a more sensitive indicator than the h-index, neither does the g-index. Taken together, g and h present a concise picture of a scientist's achievements in terms of publications and citations.

Acknowledgement. Research for this article was supported through NSFC grant 70373055.

References

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

Bar-Ilan, J. (2006). H-index for Price medalists revisited. *ISSI Newsletter* 2(1), p. 3-5.

Batista, P.D., Campiteli, M.G., Kinouchi, O. and Martinez, A.S. (2005). Universal behavior of a research productivity index. Preprint.

Braun T., Glänzel, W. ,& Schubert A.(2005). A Hirsch-type index for journals. *The Scientist*, 19(22), p.8.

Cronin, B. and Meho, L. I. (2006). Using the h-index to rank influential information scientists. *Journal of the American Society for Information Science and Technology,* 57, (to appear).

Egghe, L. (2006a). How to improve the h-index. *The Scientist*, 20(3), p. 14.

Egghe, L. (2006b). An improvement of the H-index: the G-index. *ISSI Newsletter*, 2(1), 8-9.

Egghe, L. (2006c). Dynamic h-index: the Hirsch index in function of time. *Journal of the American Society for Information Science and Technology* (to appear).

Egghe, L. and Rousseau, R. (2006). An informetric model for the Hirsch index. *Scientometrics* (to appear).

Egghe, L., Rousseau, R. and Rousseau, S. (2006). TOP-curves (Submitted to *Journal of the American Society for Information Science and Technology*).

Glänzel, W. (2006). On the h-index – a mathematical approach to a new measure of publication activity and citation impact. *Scientometrics*, 67, 315-321.

Glänzel, W. and Persson, O. (2005). H-index for Price medalists. *ISSI Newsletter*, 1(4), 5-18.

Hirsch, J.E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the USA*, 102(46), 16569-16572.

Jin, BH (2006). H-index: an evaluation indicator proposed by scientist. *Science Focus*, 1(1), 8-9.(In Chinese)

Liang, L. (2006). H-index sequence and h-index matrix: constructions and applications. *Scientometrics*, 69(1) (to appear).

Rousseau, R. (2006). Simple models and the corresponding h- and g-index. *Science Technology Development* (to appear); see also E-LIS: code ID 6153**.**