AN IMPROVEMENT OF THE H-INDEX: THE G-INDEX[1]

by L. Egghe

Universiteit Hasselt, Campus Diepenbeek, Agoralaan, B-3590 Diepenbeek, Belgium
and
Universiteit Antwerpen, Campus Drie Eiken, Universiteitsplein 1, B-2610 Wilrijk, Belgium
e-mail: leo.egghe@uhasselt.be

For a set of papers, ranked in decreasing order of the number of citations that they received, the h-index is the (unique) highest number of papers that received h or more citations. In the references [1,2,7,9] one describes some advantages of this new scientometric indicator: It is a simple single number incorporating both publication (quantitiy) and citation (quality or visibility) scores and hence has an advantage over these single separate measures and over measures such as "number of significant papers" (which is arbitrary) or "number of citations to each of the (say) q most cited papers" (which again is not a single number). The h-index is also robust in the sense that it is insensitive to an accidental set of uncited (or lowly cited) papers and also to one or several outstandingly highly cited papers.

This last point is the subject of my criticism on this measure: although I certainly agree that the insensitivity to the "tail" of lowly cited papers is an advantage for the h-index, it should be sensitive to the level of the highly cited papers. Indeed, as the h-index is defined now, once an article belongs to the h top class (defining h) it is totally unimportant whether or not these papers continue to be cited or not and, if cited, it is unimportant whether these papers receive 10, 100 or 1000 more citations! We feel that a measure which should indicate the overall quality of a scientist or of a journal should deal with the performance of the top articles and hence their number of citations should be counted, even when they are declared to be in the top class. This can be accomplished by modifying the h-index a little bit (called the g-index) so that the above described disadvantage has disappeared while keeping all advantages of the h-index and, at the same time, the calculation of the new index is as simple as the one of the h-index.

Note that it is a consequence of the definition of the h-index that the top-h papers have at least h^2 citations but that the actual number can be much higher (this is what is missing in the h-index). We therefore define the g-index as the highest number g of papers that together received g^2 or more citations. From this definition it is already clear that $g \geqq h$. So for all authors or journals, the g-score will be higher than the h-score but, what is interesting in this, the higher the number of citations in the top-class (in other words, the skewer the citation distribution) the higher the g-score will be. Let us give two real author examples: the comparison of L. Egghe and H. Small. In the Tables below, TC denotes the total number of citations to a paper on rank r and Σ TC denotes the cumulative TC scores up to rank r.

[1] This text is based on the article [3], to be published in Scientometrics.

TABLE A : L. Egghe data

TC	r	Σ TC	r^2
47	1	47	1
42	2	89	4
37	3	126	9
36	4	162	16
21	5	183	25
18	6	201	36
17	7	218	49
16	8	234	64
16	9	250	81
16	10	266	100
15	11	281	121
13	12	294	144
13	**13**	307	169
13	14	320	196
13	15	333	225
12	16	345	256
12	17	357	289
12	18	369	324
12	**19**	**381**	**361**
11	20	392	400
.	.	.	.

TABLE B : H. Small data

TC	r	Σ TC	r^2
305	1	305	1
239	2	544	4
127	3	671	9
109	4	780	16
86	5	866	25
80	6	946	36
77	7	1023	49
75	8	1098	64
67	9	1165	81
49	10	1214	100
44	11	1258	121
36	12	1294	144
26	13	1320	169
26	14	1346	196
25	15	1371	225
22	16	1393	256
22	17	1415	289

18	**18**	1433	324
18	19	1451	361
15	20	1466	400
12	21	1478	441
10	22	1488	484
9	23	1497	529
8	24	1505	576
8	25	1513	625
7	26	1520	676
6	27	1526	729
5	28	1531	784
5	29	1536	841
5	30	1541	900
3	31	1544	961
3	32	1547	1024
2	33	1549	1089
2	34	1551	1156
2	35	1553	1225
1	36	1554	1296
1	37	1555	1369
1	38	1556	1444
1	**39**	**1557**	**1521**
1	40	1558	1600
.	.	.	.

The bold face numbers indicate how the h-index and g-index is calculated. L. Egghe has h=13 since this is the last rank where all the papers have at least 13 citations. For H. Small this is h=18, higher but not so high as one would expect from the citation data of the highest cited papers of both authors. But L. Egghe has g=19 since this is the last rank for which Σ TC $\geqq g^2$. For H. Small this is g=39. Hence the difference between L. Egghe and H. Small becomes more apparent using the g-index than with the h-index. In general, in a group of authors (say of the same field) the variance of the g-indexes will be much higher than the one of the h-indexes which makes a comparison between authors concerning their visibility in the world more apparent.

Both indexes are simple to calculate based on the same table of data. We therefore hope that this new g-index will be further studied and used in practical assessments.

For a thorough study of the g-index, incl. the scores of the active De Solla Price winners we refer to [3].

In [6,8] a formula for the h-index is presented in case the data follow a Lotka power law with exponent α in the denominator. The formula is

$$h = T^{\frac{1}{\alpha}}$$

where T denotes the total number of articles. In [3] the analogous formula for the g-index has been proved to be

$$g = \left(\frac{\alpha - 1}{\alpha - 2} \right)^{\frac{\alpha - 1}{\alpha}} T^{\frac{1}{\alpha}}$$

In [4,5] a theory is presented to calculate the evolution of the h- and g-index in function of time.

REFERENCES

[1] Ball P: **"Index aims for fair ranking of scientists,".** Nature 2005, **436**:900.

[2] Braun T, Glänzel W, Schubert A: **"A Hirsch-type index for journals,".** The Scientist 2005, **19(22)**, 8-10.

[3] Egghe L: **"Theory and practise of the g-index"**. Scientometrics, to appear.

[4] Egghe L: **"Dynamic h-index: the Hirsch index in function of time"**. Journal of the American Society for Information Science and Technology, to appear, 2006.

[5] Egghe L: **"Time-dependent Lotkaian informetrics and applications to the time-dependent g-index"**. To appear, 2006.

[6] Egghe L, Rousseau R: **"An informetric model for the Hirsch-index"**. Scientometrics, to appear.

[7] Glänzel W: **"On the opportunities and limitations of the H-index,".** Science Focus 2006, **1**, in press.

[8] Glänzel W: **"On the H-index – A mathematical approach to a new measure of publication activity and citation impact".** Scientometrics **67(2)**, to appear.

[9] Hirsch JE: **"An index to quantify an individual's scientific research output,".** Proceedings of the National Academy of Science. Nov 15 2005, **102(46)**, 16569-16572. [http://xxx.arxiv.org/abs/physics/0508025].