Bibliometric Profiling of a Group: A Discussion on Different Indicators

Mobyen Uddin Ahmed and Shahina Begum
{firstname.lastname}@mdh.se
School of Innovation, Design and Engineering
Mälardalen University, PO Box 883 SE-721 23, Västerås, Sweden

Abstract. Now-a-days in some advanced countries bibliometric profiling plays a vital role when making decision on promotion, fund allocation and award prizes. Accurate identification of this is important since it is becoming important to assess scientific output for a researcher or a group of researcher. This paper presents and discusses several most common indicators of bibliometric profiling together with h- and g-indexes. A case study has been conducted on 101 scientific articles with three most well known search engines. The study results using several indicators are presented in this report.

1. Introduction

Today, a most common interest is to assess scientific productivity and apparent scientific impact that is the research performance of a scientist or a group of scientist to support research assessment decisions. This can be practiced for most of the advanced countries in their funding policies or allocation processes by the government or funding agencies. For example, the UK government is considering a process i.e. bibliometrics which shows the quality of the research output of UK universities [3]. On the basis of the assessment the UK government is allocated research funding to the universities. There are several ways to judge the performance of a scientist or group of scientists. Several widely accepted bibliometric's indicators used in the evaluation process are presented in [4], [5], [6]. However, there are different opinions about which method/indicator(s) is the best for the assessment of research performance of individual/group of scientists [7], [8]. Instead of based on a single indicator, a combination of multiple indicators is strongly recommended by different authors [9], [10].

This paper presents several well known indicators used in the process of bibliometric profiling. The main goal of this paper is to analyse the indicators together with h- and g-indexes. Moreover, difference between h- and g-index and their ability to assess research performance of a group of scientists are focused here. A case study on a research group at the School of Innovation, Design and Engineering, Mälardalen University, Västerås, Sweden has been done. Interesting observation from the results have been achieved and discussed latter in this paper.

2. Material and methods

Scientific publication of a research group (including PhD students), at the School of Innovation, Design and Engineering, Mälardalen University, Västerås, Sweden were considered for this study. The publications between the year 1995 and 2009 were collected for the 5 researchers (3

seniors and 2 PhD students) of the group. The publication information is taken from the ISS[1] home page. About one hundred articles were collected which includes book chapters, journals, conferences and workshops. Publications which have at least one citation by Google Scholar and written by at least one of the 5 researchers were considered for this study. The information about the publications is summarized in Table 1. It can be seen from Table 1, about one hundred articles have been produced in the last 15 years and around 60% of the articles have at least one citation by Google Scholar. Only 57 of the articles were considered to be calculated by several indicators. Appendix 1, 2 and 3 presents three matrixes with 57 articles and their publications information (i.e. year, author's name, and place of the publication).

Table. 1. Sample data distribution of the selected publications

Criteria	Values
Duration of year	1995-2009
Total number of years	15
Number of researchers	5
Total number of articles	101
Journals	16
Book chapters	3
Conferences	22
Workshops	16
Total number of articles (have at least one citation by the Google Scholar)	57 (≈ 60%)

The number of citations and citation rate per year for each publication are also presented in the appendix. The research performance of the group has been analyzed through a bibliometric profiling composed of the following indicators:

- Total number of article; considering journals, book chapters, conferences and workshops articles published between the period 1995 and 2009 and should have at least one citation by the Google Scholar.
- Total number of citations; three web search engines i.e. Google Scholar[2], Web of Science[3] and Scopus[4] have been used. Self-citations were included.
- Citation rate per year; the average number of citations per year is calculated for each article using all the three search engines.

[1] Intelligent Sensor System (ISS), http://www.iss.mdh.se/index.php?choice=publications&year=any&research_group=is
[2] http://scholar.google.com/
[3] http://apps.isiknowledge.com/
[4] http://www.scopus.com/

- Number of Highly Cited Articles (NHCA); the articles those have at least 5 citations using all these search engines.
- h-index, by Hirsch [1]; to quantify performance of the researchers in the group considering both the number of publications and the number of citations.
- g-index, by Egghe [2]; same as h-index but gives a higher value, a higher g-index can be achieved by the influence of a "big hit" (i.e. high number of citations).

2.1 Calculation of h-index and g-index

In order to calculate h-index and/or g-index of any group of scientists, the easiest way is to make a matrix as shown in Table 2 and find out the citation number for each article. The 1st column of the table contains ranked article number sorted in descending order based on their citation number. The 2nd column shows the citation number of each article. Now, observing these two columns values, point a position where the citation number is equal or less than the ranked article number. This pointing position of the ranked article represents the h-index number. Similarly, the 3rd and 4th columns present the square of the ranked article number and the summation of the citation numbers. Now, the g-index number is calculated by pointing a position where the square of the ranked article number is equal or greater than the accumulated number of citations. If the number of articles is not enough to calculate the g-index, the a few articles with 0 citations can be added into the matrix in order to complete the calculation. In Table 2, one example of the h-index is 6 and g-index is 10.

Table. 2. Example of h- and g-index calculation

Ranked Article	Number of Citation	Square of the Ranked Article	Summation of the Citation	
1	34	1	34	
2	18	4	52	
3	11	9	63	
4	9	16	72	
5	7	25	79	
6	6	36	85	h-index = 6
7	1	49	86	
8	1	64	87	
9	1	81	88	
10	1	100	89	g-index = 10
11	1	121	90	

3. Results

The results using several indicators as descried earlier in the 'Material and methods' are presented in Table 3 to Table 6.

Table. 3. Several Indicators using the Google Scholar

Criteria		Values
Total number of citation		517
Maximum number of citation for a paper		134
Highly cited paper (>=5 citation)		24
Average citation for the 57 articles		9
h-index		11
g-index		21
The ratio of the g- and h-indexes		1.9
Citation rate per year and per article	Maximum	17
	Minimum	0.08
	Average	1.6
	Standard deviation	2.5
	Median	1

In Table 3, citations by Google Scholar were considered, where the number of *total citation* is found as 517 for the 57 articles. There are 24 articles which classified as *highly cited* article, i.e. more than 5 citations. The *maximum number of citation* is found as 134 for one article. The h- and g-index are calculated as 11 and 21. Moreover, calculated statistics for the citation rate per year for each of the article are also presented in Table 3. The same indicators were used for the Web of Science and Scopus and the results are illustrated in Table 4 and Table 5.

Table. 4. Several Indicators using Web of Science

Criteria		Values
Maximum citation		17
Total citation		61
High citation paper (>=5)		4
Average citation on 19 articles indexed by the Web of Science		3.2
h-index		5
g-index		8
The ratio of g- and h-indexes		1.6
Citation rate per year and per article	Maximum	2.2
	Minimum	0.0
	Average	0.2
	Standard deviation	0.5
	Median	0

Table. 5. Several Indicators using Scopus

Criteria		Values
Maximum citation		70
Total citation		173
High citation paper (>=5)		10
Average citation on 20 articles indexed by the Scopus		8.7
h-index		6
g-index		10
The ratio of g- and h-indexes		1.7
Citation rate per year and per article	Maximum	8.8
	Minimum	0.0
	Average	0.5
	Standard deviation	1.3
	Median	0

A comparison of the articles published within 1999-2005 is presented in Table 6 with a breakdown on every 5 years. Here only Google Scholar is considered for the comparison and present the number of articles published in Web of Science and Scopus in that duration.

Table. 6. Comparison of the articles published within 1999-2005 (with a breakdown on every 5 years)

Duration of Years	h-index	g-index	g/h	Total Number of Article	Total Number of Citation	Number of Articles in Web of Science	Number of Articles in Scopus
2005-2009	5	8	1.6	27	103	9	12
2004-2000	9	19	2.1	21	365	8	7
1999-1995	4	7	1.75	9	49	2	1

4. Discussion

The production of the researchers of on the research group at Mälardalen University is amounted to 101 articles during 1995-2009 as shown in Table 1. To extract information we used three well known databases i.e. Scopus, Web of Science, and Google Scholar. Google scholar originates from the USA and other two from the Europe [11]. The citation analysis done by these three search engines are presented in Table 3, Table 4 and Table 5. The indicators used are generally based on the no. of production as well as on the impact of the publication i.e. no. of citations, no. of citation per document, no. of highly cited paper. The h-index measures the quantity and number of citation i.e. impact of a scholarly literature. Highly cited papers are important for h-index, however, number of citations is unimportant here. So to overcome this problem g-index is introduced.

From Table 3, Table 4 and Table 5 it can be observed that the maximum number of citation using Scopus and Web of Science is 70 and 17 respectively whereas with Google Scholar it is 134. So it shows a much higher values compare with the other two. Also in most of the other

indicators Google Scholar provides higher values compare to the Web of Science and Scopus. The different search criteria for different databases could bias the result. Google Scholar is free accessible Web search engine, it retrieves all the possible electronic references whereas, Scopus and Web of Science are commercial databases so they require access fees. The latter two updates for printed version of the literature, they don't consider any online early version. However, the values of h- and g-index and their ratio (Web of Science=5, 8, and 1.6; Scopus=6, 10, and 1.7) are close for the both Web of Science and Scopus. Moreover, average citation is almost same both for Google and Scopus, i.e. 9 and 8.7. This could be due to the fact that out of 57 articles only 19 articles are indexed by the Web of Science and 20 articles are indexed by the Scopus. There are some articles, which are indexed by the all search engines (3 mentioned above), some of them are only indexed either by Web of Science or Scopus, and the rests of them are only considered by Google Scholar. In our study, Google Scholar provides better result compare to other two search engines (Web of Science and Scopus) since it can cover a broader area of published articles and it is very easy to search.

According to the number of citation analysis by Google Scholar (see Table 6), it can be observed that the group has focused more on the quantity of the publication rather than the no. of citations during the year 2005-2009. There are more than 40 articles which have not been cited even by the authors themselves. As can be seen from Table 6, during 2000 to 2004, better h- and g-indexes are achieved compare to the other years. However, 27 articles out of 57 are produced during the year 2005 to 2009. The number of articles indexed by the Web of Science and Scopus is also more in the year between 2005 and 2009.

It is also observed that *number of citation* is not a good indicator since it does not consider publication year of an article. For example, researcher 'A' has an article with 50 citations in 10 years and researcher 'B' has an article with 10 citations in 1 year, then researcher 'A' will always get high ranking. On the other hand, researcher 'B' might get 50 citations in 5 years since his article has better citation rate. Therefore, *citation rate per year* might be a better indicator to assess research performance. To calculate h- and g-indexes, *number of citations* of an article is important. However, for h-index quantity of publications plays an important role. So it could influence to publish more articles instead of encouraging quality. Although, there are lot of debate on if the h-index is better than the g-index or vice versa [7], [8] but in this report we have considered also the ratio (i.e. g/h) of them. According to our results, the ratio between g- and h-indexes (i.e. h/g) shows close result using all the three search engines and this could be a good indicator to access scientific performance of a researcher.

In fact, Bibilometric profiling to access research performance of a researcher or a group of researcher is a complex task and could be accessed combining several indicators. However, in some cases it is domain dependent since different domain or research area has different practice/strategy for their publications and citations.

References

1. Hirsch, J. E. (2005), An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences of the United States of America, Vol. 102, No. 46, pp 16569–16572.
2. Egghe, L. (2006), Theory and practise of the g-index. Scientometrics, Vol. 69, No. 1, pp 131–152.
3. Higher Education Funding Council for England, http://www.hefce.ac.uk/research/ref/Biblio/. Accessed 25 September 2010.
4. Martin, B. R. (1996), The use of multiple indicators in the assessment of basic research. Scientometrics, Vol. 36, No. 3, pp 343–362.
5. Lewison, G., Cottrell, R., Dixon, D. (1999), Bibliometric indicators to assist the peer review process in grant decisions. Research Evaluation, Vol. 8, No.1, pp 47–52.
6. Van Leeuwen, T. N., Visser, M. S., Moed, H. F., Nederhor, T. J., Van Raan, A. F. J. (2003), The Holy Grail of science policy: Exploring and combining bibliometric tools in search of scientific excellence. Scientometrics, Vol. 57, No. 2, pp 257–280.
7. Rodrigo C. & Maria B., (2008), Is g-index better than h-index? An exploratory study at the individual level, Scientometrics, Vol. 77, No. 2, pp 267–288.
8. Rodrigo C. & Maria B., (2007), The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level, Journal of Informetrics, Vol. 1, No. 3, pp. 193-203.
9. Van Raan, A. F. J. (2006), Comparisons of the Hirsch-index with standard bibliometric indicators and with peer judment for 147 chemistry research groups. Scientometrics, Vol. 67, No. 3, pp 491–502.
10. Glanzel, W. (2006), On the h-index – A mathematical approach to a new measure of publication activity and citation impact. Scientometrics, Vol. 67, No. 2, pp 315–321.
11. Matthew E. F., Eleni I. P., George A. M., and Georgios P., (2008), Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. The FASEB Journal. Vol. 22, No. 2, pp 338-342.

Additional useful links

1. The h Index for Computer Science, http://www.cs.ucla.edu/~palsberg/h-number.html, Accessed 25 September 2010.
2. Journal ranking on h-index, http://www.scimagojr.com/journalrank.php, Accessed 25 September 2010.
3. Country ranking on h-index, http://www.scimagojr.com/countryrank.php, Accessed 25 September 2010.