

An outgrow index

Ronald Rousseau^{1,2,3} and Xiaojun Hu^{4,5}

¹ KHBO (Association K.U.Leuven), Industrial Sciences and Technology, Zeedijk 101, B-8400 Oostende, Belgium

²University of Antwerp, IBW, Venusstraat 35, B-2000 Antwerpen, Belgium

³ K.U.Leuven, Dept. Mathematics, Celestijnenlaan 200B, B-3000 Leuven (Heverlee), Belgium, E-mail: ronald.rousseau@khbo.be

⁴College of Public Administration, Zhejiang University, Hangzhou, China

⁵Medical Information Centre, Zhejiang University School of Medicine, Hangzhou, China, E-mail: xjhu@zju.edu.cn

Proposes a relative index measuring the amount by which an article outgrows, in terms of citations, the publications on which it is based. The study involves citations collected from Web of Science during the last week of April 2010 along with the number of citations received (also in WoS) by each of the references.

Introduction

We define citation analysis as that subfield of informetrics where patterns and frequencies of citations, given as well as received are analyzed. Such an analysis is performed on the level of authors, journals, scientific disciplines and any other useful unit or level. Citation analysis further studies relations between cited and citing units (documents, authors, countries etc.). From an application point of view citation analysis may be considered as a collaborative peer effort to analyze and promote the quality of scholarly publication and research^{1,2}.

When Garfield started collecting information from bibliographic data and references in scientific papers in the early 1960s, he could not imagine the storms he would raise in the realms of research evaluation. Nowadays, scientometric indicators derived from citation data provided by Thomson Reuters' databases or one of its competitors and discussions about their meaning and usefulness pervade academia. The discussions related to the h-index^{3,4}, the crown indicator^{5,6,7} and the UK Research Assessment Exercise^{8,9} are recent examples of this phenomenon. As such one may rightly say that Garfield created a *Butterfly Effect*^{10,11}.

The research question tackled in this contribution is to define an index describing to which extent an article outgrows its own references in terms of received citations.

To do this, we stand on the shoulders of Eugene Garfield's life work (actually a part of, as his realizations go far beyond the one used here), namely the Web of Science (WoS).

Methodology

We consider articles, mainly published in journals included in the JCR subject category *Medical, General & Internal*, collected the number of received citations, with and without author self-citations (data were collected from the WoS, during the last week of April 2010) and also collected the number of citations received (also in the WoS) by each of its reference items. In case the reference item was included in the WoS we just recorded the number of received citations as provided by the article's bibliographic record. If not we performed a Cited Reference Search and collected all citations that seem to match with the retrieved item. It is well-known that spelling variants and incomplete bibliographic information makes it impossible to determine the exact number of received citations yet we tried to obtain a reasonable approximation (except for the case of a book which was clearly more cited than all other publications involved in the calculation of an outgrow index). We also noted if a reference item was a book or not.

We consider articles that are at least ten years old leading to a stable situation. Yet, it might be interesting to study a time series of outgrow indices starting from the year of

publication. We discuss this point in the final section of this article.

Definition: The outgrow index

If the reference list of article A contains R items then we rank these according to the number of citations received. Adding article A to this list, we obtain a list of R+1 items, ranked according to the number of citations received. In case of equal number of received citations, we assign the average rank. If article A ranks R_A -th place then its index is

$$I(A) = 1 - \frac{R_A}{R + 1}$$

This index takes values between 0 (included) and 1 (not included). We do not include the value 1, i.e. we do not define I(A) as $1 - \frac{R_A}{R}$, where rankings begin at 0. This choice is made to express that it is easier to obtain the highest number of citations among a small number of articles, than among a large number of articles. Hence receiving more citations than one's nine references leads to an outgrow index of $1 - 1/10 = 0.9$, while receiving more citations than one's 49 references leads to an outgrow index of $1 - 1/50 = 0.98$.

Besides the general index we propose some variants, some of which are used in the examples we provide:

1. Removing self-references, leading to a reference list of smaller length (or of the same length).
2. Removing books from the reference list, as it is usually very difficult for a research article to obtain more citations than a standard reference book.
3. Removing all reference items that are not WoS source items (such as books) leading to a WoS outgrow index.

Moreover, citations received by A can be calculated including author self-citations or not.

Results

For illustrative purposes we mainly collected data for highly-cited articles in medicine, but included also a few other ones. As a first example we propose a classic in our own field, namely Henry Small's article introducing the notion of co-citation.

Title: cocitation in scientific literature - new measure of relationship between 2 documents

Author: Small H

Source: Journal of the American Society for Information Science, Volume 24, Issue 4, Pages 265-269,

Published: 1973

Times Cited: 474 (of which 20 are self-citations)

This article has 10 references, which are cited (ranked in decreasing order and separated by a semi-colon): 1,454 ; 448 ; 433 ; 273 ; 40 ; 30 ; 28 ; 27 ; 11 ; 5 times. Hence, its outgrow index is $1 - 2/11 = 0.818$. Removing self-citations does not change this article's rank (474 becomes 454, which is still the second most cited article). Three publications (one book and two articles in congress proceedings) are not included as WoS sources. Hence Small's WoS outgrow index is $1 - 2/8 = 0.75$. We note that the articles analyzed in Small's publication are given in a separate table. However, the WoS has included these as if they were normal references. Using the reference list as given in the WoS we have 18 references. The added references are all highly cited so that Small's article becomes the 10th in the ranked list according to citations. In this way his outgrow index is only $1 - 10/19 = 0.474$.

Some examples of highly cited articles in medicine

Title: Statistical-methods for assessing agreement between 2 methods of clinical measurement

Author(s): Bland JM, Altman DG

Source: Lancet Volume: 327 Issue: 8476 Pages: 307-310, Published:1986

Times cited: 16,906 (16,868 without self-citations)

This article contains eight references, which are cited: 1,195; 130 ; 41 ; 21 ; 12 ; 3 ; 3 ; 1 times. Consequently this article's outgrow index is $1 - 1/9 = 0.889$. Removing self-citations does not make a difference to this article's outgrow index. As four of these references are not source articles, the WoS outgrow index is $1 - 1/5 = 0.8$.

Title: Assessment of outcome after severe brain-damage - practical scale

Author(s): Jennett B, Bond M

Source: Lancet Volume: 305 Issue: 7905 Pages: 480-484, Published: 1975

Times Cited: 3,259 (3,216 without self-citations)

This article contains 24 references, which are cited 4350; 460 ; 151 ; 112 ; 105 ; 96 ; 89 ; 77 ; 72; 67 ; 65 ; 59 ; 53; 29; 27; 23 ; 12 ; 6 ; 5 ; 4 ; 2 ; 1 ; 1 ; 1 times. Consequently this article's outgrow index is $1 - 2/25 = 0.92$. Removing self-citations does not make a difference. As 13 of the references (not the most-cited one) are not source articles, the WoS outgrow index is $1 - 2/12 = 0.833$.

Title: unidentified curved bacilli in the stomach of patients with gastritis and peptic-ulceration
 Author(s): Marshall BJ, Warren JR
 Source: lancet Volume: 323 Issue: 8390 Pages: 1311-1315, Published: 1984
 Times Cited: 2,359 (including one self-citation)

This articles contains 29 references, which are cited 3,621; 2,594; 1,395; 753; 565; 557; 557; 458; 388; 326; 319; 239; 226; 158; 142; 136; 121; 110; 102; 94; 64; 52; 38; 30; 29; 16, 6; 2; 1 times. Consequently this article's outgrow index is $1 - 3/30 = 0.9$. Removing self-citations does not make a difference. As 8 of the references (not the top 2) are not WoS source articles, the WoS outgrow index is $1 - 3/22 = 0.864$.

Title: a new simplified acute physiology score (saps-ii) based on a European North-American multicenter study
 Author: Legall JR, Lemeshow S, Saulnier F
 Source: JAMA-Journal of the American Medical Association, Volume 70, Issue 24, Pages 957-2963, Published 1993
 Times Cited: 1,805 (including three self-citations)

This article contains 21 references which are cited 5,849; 5,791; 2,732; 1,555; 914; 841; 732; 698; 558; 449; 221; 202; 185; 156; 114; 82; 66; 62; 43; 7; 3 times. Consequently this article's outgrow index is $1 - 4/22 = 0.818$. Again, removing self-citations does not make a difference. As 5 of the references (not the top 3) are not WoS source articles, the WoS outgrow index is $1 - 4/17 = 0.765$.

An example of an article with a short reference list

Title: Prematurity and uniqueness in scientific discovery
 author(s): Stent GS

Source: Scientific American Volume: 227 Issue: 6
 Pages: 84-&
 Published: 1972, Times Cited: 120 (including one self-citation)

This article contains three references:

Churchman, C.W. (1966). Hansel CEM – ESP – A scientific evaluation. *Science*, 153 (3740), pp.1088-. Cited 3 times.

Polanyi, M. (1963). Potential theory of adsorption. *Science*, 141(358), pp. 1010-. Cited 50 times.

Stent, G.S. (1971). *Molecular Genetics. An introductory narrative*. San Francisco: W.H. Freeman. This is a book, hence not a WoS source item: cited 60 times.

As this article's citations (with or without self-citations) is larger than its references' its outgrown index is at a maximum, namely $1 - 1/4 = 0.75$. Removing books, or removing non-WoS items (which amounts to the same here) reduces the outgrown index to 0.667. This example illustrates the fact that it is relatively easy to outgrow a short reference list. This case is an example where the reference list includes a book that has received a fair number of citations but not as much as the article for which we determine an outgrow index.

Finally we also add an article (one of our own) that did not outgrow its reference list at all.

Title: A characterization of distributions which satisfy price law and consequences for the laws of zipf and mandelbrot

Author(s): Egghe I, Rousseau R

Source: Journal of Information Science Volume: 12
 Issue: 4 Pages: 193-197

Published: 1986

Times Cited: 6 (two self-citations, by Egghe)

This article contains 9 references. However, two relate to (at that time) unpublished items, which we do not take into account. This leaves 7 references which are cited as follows: (> 600); 328 ; 45 ; 28 ; 18 ; 17 ; 1. Consequently this article's outgrow index is $1 - 7/8 = 0.125$. Of these two, the most-cited (a book) and the least cited (an item in an encyclopaedia) are not WoS source items. Hence the article's WoS outgrow index is: $1 - 6/6 = 0$. Taking self-citations into account or not makes no difference.

We note that, as we study references of a WoS publication all reference items are cited at least once (by construction there are no non-cited items). Yet, it is interesting to note that some of the reference items in our examples are cited exactly once.

Conclusion and comments

This contribution was written somewhat in the spirit of Gene Garfield's Essays of an Information Scientist. We did not write a full-fledged scientific article but, hopefully, proposed an interesting indicator based on citation data.

The outgrow index is somewhat related to the Ortega Hypothesis, but our aim is much more modest. We just want to find a nice description or representation. Yet, our outgrow index might be used for testing the Ortega Hypothesis. Recall that the so-called *Ortega Hypothesis* states that the work of the scientific elite owes much to the work of the average scientist, and that the work of great scientists is built upon a pyramid of small discoveries made by these scientists^{12,13,14,15}.

Based on the few examples studied it seems that taking self-citations into account or not does not make a lot of difference. Not surprisingly, highly cited articles easily outgrow their reference list. Clearly, much more outgrow indices must be determined before it becomes possible to state general conclusions.

Studying the change of outgrow indices over time, according to one of the many possible time series¹⁶, would be an interesting next step.

Acknowledgement

Xiaojun Hu was supported by a grant from the Fundamental Research Funds for the Central Universities.

References

1. Spinak E, *Diccionario Enciclopédico de Bibliometría, Cienciométrica e Informetría*: Caracas: UNESCO-CII/II. 1996.
2. Rousseau R, Publication and citation analysis as a tool for information retrieval. In: *Social Information Retrieval Systems. Emerging technologies and applications for searching the web effectively* (Goh D & Foo S, eds.). Hershey (PA); Information Science Reference (IGI Global). 2008. Chapter 13; pp. 252-267.
3. 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*, 102 (2005) 16569-16572.
4. Waltman L, Van Eck N J, A simple alternative to the *h*-index, *ISSI Newsletter*, 5(3) (2009) 46-48.
5. Lundberg J, Lifting the crown—citation *z*-score, *Journal of Informetrics*, 1(2) (2007) 145-154.
6. Opthof T, Leydesdorff L, Caveats for the journal and field normalizations in the CWTS ("Leiden") evaluations of research performance, *Journal of Informetrics*, 4(3) (2010) 423-430.
7. Waltman L, Van Eck N J, Van Leeuwen T N, Visser M S, Van Raan A F J, Towards a new crown indicator: Some theoretical considerations. (2010) arXiv:1003.2167.
8. Warner J, A critical review of the application of citation studies to the Research Assessment Exercises, *Journal of Information Science*, 26(6) (2000) 453-459.
9. Oppenheim C, Out with the old and in with the new: the RAE, bibliometrics and the new REF, *Journal of Librarianship and Information Science*, 40(3) (2008) 147-149.
10. Lorenz E N, Deterministic nonperiodic flow, *Journal of the Atmospheric Sciences*, 20(2) (1963) 130-141.
11. Gleick J, *Chaos. Making a new science*: New York: Viking. 1987.
12. Cole J R, *The social structure of science: a study of the reward and communications systems of modern physics*. Ph.D. dissertation: Columbia University. 1969.
13. Cole J R, Cole S, The Ortega hypothesis: citation analysis suggests that only a few scientists contribute to scientific progress, *Science*, 178 (1972) 368-375.
14. Hoerman H L, Nowicke C E, Secondary and tertiary citing: a study of referencing behavior in the literature of citation analysis deriving from the Ortega hypothesis of Cole and Cole, *Library Quarterly*, 65(4) (1995) 415-434.
15. Rousseau R, Yang L Y, Yue T, A discussion of Prathap's h_2 -index for institutional evaluation with an application in the field of HIV infection and therapy, *Journal of Informetrics*, 4(2) (2010) 175-184.
16. Liu Y X, Rousseau R, Definitions of time series in citation analysis with special attention to the *h*-index, *Journal of Informetrics*, 2(3) (2008) 202-210.