BRIEF COMMUNICATION

Dynamic *h*-Index: The Hirsch Index in Function of Time

Leo Egghe
Universiteit Hasselt (UHasselt), Campus Diepenbeek, Agoralaan, B-3590 Diepenbeek, Belgium[*] and Universiteit Antwerpen (UA), Campus Drie Eiken, Universiteitsplein 1, B-2610 Wilrijk, Belgium. E-mail: leo.egghe@uhasselt.be

When there are a group of articles and the present time is fixed we can determine the unique number h being the number of articles that received h or more citations while the other articles received a number of citations which is not larger than h. In this article, the time dependence of the h-index is determined. This is important to describe the expected career evolution of a scientist's work or of a journal's production in a fixed year. We use the earlier established cumulative n^{th} citation distribution. We show that

$$h = ((1 - a^t)^{\alpha-1} T)^{\frac{1}{\alpha}}$$

where a is the aging rate, α is the exponent of Lotka's law of the system, and T is the total number of articles in the group. For $t = +\infty$ we refind the steady state (static) formula $h = T^{\frac{1}{\alpha}}$, which we proved in a previous article. Functional properties of the above formula are proven. Among several results we show (for α, a, T fixed) that h is a concavely increasing function of time, asymptotically bounded by $T^{\frac{1}{\alpha}}$.

Introduction

Let us take any "group" of articles: This can be the list of publications of a scientist, the set of articles in a year of publication in a journal, a bibliography, and so on. The h-index is this number such that we have exactly h papers with h or more citations while the other articles have not more than h citations. This h-index has been introduced in Hirsch (2005; see also Ball, 2005). Introduced in physics, the h-index was also well received by informetricians—see Bornmann and Daniel (2005) and Braun, Glänzel, and Schubert (2005). The h-index has the following advantages (see Hirsch, 2005, Braun, Glänzel, and Schubert, 2005):

- It is a single number incorporating both publication and citation scores and hence has an advantage over these single measures and over measures such as "number of significant articles" (which is arbitrary—and is not so for the h-index) or "number of citations to each of the q most-cited articles" (which again is not a single number).
- It is robust in the sense that it is insensitive to an accidental set of uncited (or lowly cited) articles and also to one or several outstandingly highly cited articles.
- It combines the features quantity (number of articles) with quality (or visibility) in the sense of citation rates.

I share the opinion of Braun et al. (2005) that the h-index will be the topic of many informetrics articles in the (near) future and hence, it deserves a concise mathematical study. In Egghe and Rousseau (2006) we have already shown that any system has a unique h-index (there the h-index notion was extended to general information production processes, where sources, e.g., articles, produce items, e.g., citations). Further, we showed that, when we have a Lotkaian system, i.e., a system such that the law of Lotka is valid (see Egghe & Rousseau, 1990, or Egghe, 2005).

$$f(j) = \frac{C}{j^{\alpha}} \tag{1}$$

$C > 0$, $\alpha > 1$, $j \geq 1$, and where we have T sources (e.g., articles) in total, the h-index equals

$$h = T^{\frac{1}{\alpha}} \tag{2}$$

Properties of this functional relation are also proved in Egghe and Rousseau (2006).

Here we are interested in the evolution of the h-index over time for a set of articles. In Egghe and Rao (2001) we determined the cumulative n^{th} citation distribution, i.e., the cumulative distribution $\Gamma_n(t)$ of the times t at which the articles (in such a general set of articles) receive their n^{th} citation, in other words: the cumulative fraction of articles (among the ever-cited articles) that have received n citations at time t. In this article, we show that such a distribution can be used to calculate, for every t, the (hence t-dependent) h-index.

[*]Permanent address.

Received January 23, 2005; accepted February 15, 2006

 • Published online 27 December 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.20473

In the next section, we will elaborate the model, where we will prove that the time dependent h-index equals

$$h = ((1 - a^t)^{\alpha - 1}T)^{\frac{1}{\alpha}} \tag{3}$$

where T denotes the total number of ever-cited articles, α is Lotka's exponent (the system articles–citations supposed to be Lotkaian) and where a is the aging rate of the citations.

We then prove the following properties of Equation 3:

- The h-index as function of t (a, α, T constant) is a concavely increasing function with horizontal asymptote at height $T^{\frac{1}{\alpha}}$. Hence, for t going to ∞, we refind the time-independent result, Equation 2, proved in Egghe and Rousseau (2006).
- The h-index as function of α (a, T, t constant) and a (α, T, t constant) is decreasing and (of course) increasing in T (a, α, t constant).

The *h*-Index as Function of Time *t*

Let us recall a result, proved in Egghe and Rao (2001) on the cumulative n^{th} citation distribution, i.e., the cumulative distribution over time at which an article will receive its n^{th} citation ($n = 1, 2, 3, \ldots$). Here the fractions are calculated with respect to the population of eventually cited articles. We have the following result.

Theorem II.1 (Egghe & Rao, 2001)

(i) Let $C(t)$ denote the cumulative citation distribution, i.e., the cumulative distribution of the fractions of citations at time t (i.e., time t after publication). Then the cumulative n^{th} citation distribution $\Gamma_n(t)$ (with respect to all papers that are ever cited) is equal to

$$\Gamma_n(t) = \left(\frac{C(t)}{n}\right)^{\alpha-1} \tag{4}$$

where $\alpha > 1$ denotes Lotka's exponent in the law of Lotka (1) that describes the article–citations relationship.

(ii) In case of exponential aging, with aging rate a, $0 < a < 1$, we have that $C(t) = 1 - a^t$ and hence

$$\Gamma_n(t) = \left(\frac{1 - a^t}{n}\right)^{\alpha-1} \tag{5}$$

From this, the t-dependent h-index can be determined.

Theorem II.2

The t-dependent h-index equals

$$h = h(t, \alpha, T) = (C(t)^{\alpha-1}T)^{\frac{1}{\alpha}} \tag{6}$$

for $t \geq 0$ and where T denotes the total number of ever-cited articles under study. In the special case of (ii) in the above theorem we have

$$h = h(t, \alpha, T, a) = ((1 - a^t)^{\alpha-1}T)^{\frac{1}{\alpha}} \tag{7}$$

Proof:

For every n, $\Gamma_n(t)T$ is the fraction (with respect to the ever-cited papers) of articles that have n or more citations at time t. The definition of the h-index gives that $h = n$ for this n such that

$$\Gamma_n(t)T = n. \tag{8}$$

Indeed, $\Gamma_n(t)T$ is the number of papers with n citations at time t or before, hence, the number of articles with n or more citations (and automatically the other papers have less than n citations: this is, in the continuous setting, equal to "no more than n citations"). So Equation 8 is the defining relation for the (t-dependent) h-index: $h = n$. Using Equation 4, we hence have

$$\Gamma_h(t) = \left(\frac{C(t)}{h}\right)^{\alpha-1} T = h$$

or

$$h = (C(t)^{\alpha-1}T)^{\frac{1}{\alpha}}$$

In case $C(t) = 1 - a^t$ (exponential aging rate) we evidently have Equation 7. □

The next corollary was proved in Egghe and Rousseau (2006; see also, Glänzel, 2006) for an approximation in the discrete case); hence, it also belongs to the present time-dependent theory as a limiting case.

Corollary II.3

If we let $t \to \infty$ we have, in all cases:

$$h = T^{\frac{1}{\alpha}} \tag{9}$$

Result (Equation 9) was proved in Egghe and Rousseau (2006) without supposing any aging distribution: There we only used the distribution of papers with a certain number of citations that follows Lotka's law with exponent $\alpha > 1$. In this model, the minimum number of citations is 1; hence, we were only dealing with articles that are eventually cited. This is also the case in the present model. The cumulative distribution of the time of citations is given by $C(t)$. Supposing that $\lim_{t\to\infty} C(t) = 1$ automatically implies that we only consider articles that are, eventually, cited. Therefore, it is logical that the result (Equation 9) is found here again. It is, nevertheless, remarkable that the static (time-independent) Lotka result is found here as a limiting result of our dynamic (time-dependent) theory for $t \to \infty$.

We have the following further corollaries of Theorem II.2.

Corollary II.4

The function $h(t)$ in Theorem II.2 in the variable t (while all the other variables are kept constant) is a concavely strictly increasing function that ranges between 0 and $T^{\frac{1}{\alpha}}$ (see Figure 1).

DOI: 10.1002/asi

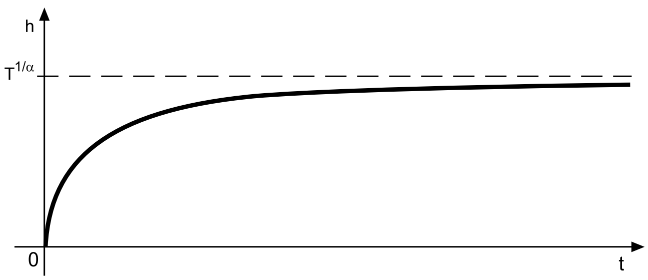

FIG. 1. Graph of the time-dependent h-index.

Proof:

It is readily seen that $h'(t) > 0$ and $h''(t) < 0$ for all $t \geq 0$. Furthermore, $h'(0) = +\infty$ and $\lim_{t\to\infty} h'(t) = 0$. Finally $h(0) = 0$ and $\lim_{t\to\infty} h(t) = T^{\frac{1}{\alpha}}$, the highest value. □

The result in Corollary II.4 is logical but shows that, when time passes, it becomes more and more difficult to increase the h-index of the set of articles under study, e.g., for a set of journal articles in a certain year or for a scientist's publications. Note that in this last case we neglect the fact that publications have different publication times. We leave it as an open problem to extend this theory to this case but the present theory is a good approximation for that, certainly for high t as discussed above.

Corollary II.5

(i) For a, T, t fixed we have that h is a decreasing function of α.
(ii) For α, T, t fixed we have that h is a decreasing function of a.
(iii) For α, a, t fixed we (evidently) have that h is an increasing function of T.

Proof:

This follows readily by calculating the corresponding derivatives. □

References

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

Bornmann, L., & Daniel, H.-D. (2005). Does the h-index for ranking of scientists really work? Scientometrics, 65(3), 391–392.

Braun, T., Glänzel, W., & Schubert, A. (2005). A Hirsch-type index for journals. The Scientist 19(22), 8–10.

Egghe, L. (2005). Power laws in the information production process: Lotkaian informetrics. Oxford: Elsevier.

Egghe, L., & Rao, I.K. Ravichandra. (2001). Theory of first-citation distributions and applications. Mathematical and Computer Modelling, 34, 81–90.

Egghe, L., & Rousseau, R. (1990). Introduction to informetrics. Quantitative Methods in Library, Documentation and Information Science. Amsterdam: Elsevier.

Egghe, L., & Rousseau, R. (2006). An informetric model for the Hirsch index. Scientometrics, 69(1), 121–129.

Glänzel, W. (2006). On the H-index—A mathematical approach to a new measure of publication activity and citation impact. Scientometrics 67(2), 315–321.

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, 16569–16572.

DOI: 10.1002/asi