Journal of Informetrics 7 (2013) 388–393

Contents lists available at SciVerse ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

A mathematical characterization of the Hirsch-index by means of minimal increments

L. Egghe[a,b,*]

[a] Universiteit Hasselt (UHasselt), Campus Diepenbeek, Agoralaan, B-3590 Diepenbeek, Belgium
[b] Universiteit Antwerpen (UA), IBW, Stadscampus, Venusstraat 35, B-2000 Antwerpen, Belgium

ARTICLE INFO

Article history:
Received 27 November 2012
Accepted 4 January 2013
Available online 14 February 2013

Keywords:
Hirsch-index
h-Index
Characterization
Increment

ABSTRACT

The minimum configuration to have a *h*-index equal to *h* is *h* papers each having *h* citations, hence h^2 citations in total. To increase the *h*-index to $h+1$ we minimally need $(h+1)^2$ citations, an increment of $I_1(h)=2h+1$. The latter number increases with 2 per unit increase of *h*. This increment of the second order is denoted $I_2(h)=2$.

If we define I_1 and I_2 for a general Hirsch configuration (say *n* papers each having $f(n)$ citations) we calculate $I_1(f)$ and $I_2(f)$ similarly as for the *h*-index. We characterize all functions *f* for which $I_2(f)=2$ and show that this can be obtained for functions $f(n)$ different from the *h*-index. We show that $f(n)=n$ (i.e. the *h*-index) if and only if $I_2(f)=2$, $f(1)=1$ and $f(2)=2$.

We give a similar characterization for the threshold index (where *n* papers have a constant number *C* of citations). Here we deal with second order increments $I_2(f)=0$.

1. Introduction

Let us have a set of *n* papers where the *i*th paper ($i=1, \ldots, n$) has c_i citations (i.e. received c_i citations). We assume that papers are arranged in decreasing order of received citations (i.e. $c_i \geq c_j$ if and only if $i \leq j$). The Hirsch-index (or *h*-index) (see Hirsch, 2005) is the highest rank $r=h$ such that all papers on ranks 1, 2, . . ., *h* have at least *h* citations, i.e. $c_i \geq h$ for $i=1, \ldots, h$ and *h* is the highest rank with this property.

In general $c_i > h$ in many cases for $i=1, \ldots, h$ but the minimum situation to have a *h*-index equal to *h* is to have *h* papers each with exactly *h* citations and where the other papers have zero citations. At least from a theoretical point of view this situation is mathematically interesting. We are interested in the minimum situation to have a *h*-index equal to $h+1$. Of course, this is a situation where we have $h+1$ papers each with exactly $h+1$ citations and where the other papers have zero citations. In the first case we require h^2 citations while in the second case we require $(h+1)^2$ citations. The increment of order 1 we define as

$$I_1(h) = (h+1)^2 - h^2 = 2h+1 \quad (1)$$

* Correspondence address: Universiteit Hasselt (UHasselt), Campus Diepenbeek, Agoralaan, B-3590 Diepenbeek, Belgium. Tel.: +32 11268121; fax: +32 11268126.
E-mail address: leo.egghe@uhasselt.be

1751-1577/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.joi.2013.01.005

The increment of order 2 we define as

$$I_2(h) = I_1(h+1) - I_1(h) \quad (2)$$

$$\begin{aligned} I_2(h) &= 2(h+1) + 1 - 2h - 1 \\ I_2(h) &= 2 \end{aligned} \quad (3)$$

So for the h-index, $I_2(h) = 2$, for every value of h.

Let us consider another example: the threshold index t (called the "highly cited publications indicator" in Waltman & van Eck, 2012). Let us fix a certain constant $C > 0$ (the threshold). Then the system has threshold index t if $r = t$ is the highest rank such that all papers on rank 1, . . ., t have at least C citations, i.e. $c_1 \geq C$ for 1, . . ., t and t is the highest rank with this property.

Again, in general $c_1 > C$ in many cases for 1, . . ., t but the minimum situation to have a threshold index equal to t is to have t papers with exactly C citations and where the other papers have zero citations. To have the minimal situation for a threshold index equal to $t+1$ we need $t+1$ papers with exactly C citations and while the other papers have zero citations. In the first case we have tC citations while in the second case we have $(t+1)C$ citations. Now the increment of order 1 is

$$I_1(t) = (t+1)C - tC = C \quad (4)$$

and hence the increment of order 2 is

$$\begin{aligned} I_2(t) &= I_1(t+1) - I_1(t) \\ I_2(t) &= 0 \end{aligned} \quad (5)$$

So for the threshold index $I_2(t) = 0$ for every value of t.

We now generalize these two situations by using a general increasing function $f(n)$ (n = 1, 2, 3, . . .). The most general h-type index definition is as follows: we have an index value n if $r = n$ is the highest rank such that all papers on ranks 1, . . ., n have at least $f(n)$ citations. Note that $n = h$ for $f(n) = n$ for each n and that $n = t$ for $f(n) = C$ (a constant as above) for each n. Other examples are: $n = h^{(2)}$ (Kosmulski's index $h^{(2)}$ – see Kosmulski, 2006) for $f(n) = n^2$ and $n = w$ (Wu's index w – see Wu, 2010) for $f(n) = 10n$ or their generalizations (see Egghe, 2011): generalized Kosmulski-index $h^{(a)}$ for $f(n) = n^a$ and generalized Wu-index w_a for $f(n) = an$, where $a > 0$ is a constant.

Again, in general, $c_i > f(n)$ in many cases for i = 1, . . ., n but the minimum situation to have an index equal to n is to have n papers with exactly $f(n)$ citations and where the other papers have zero citations. In this case we have a total of $nf(n)$ citations. To have the minimal situation for an index equal to $n+1$ we need $n+1$ papers with exactly $f(n+1)$ citations. Now we have a total of $(n+1)f(n+1)$ citations. Now the general increment of order 1 is, for every n,

$$I_1(n) = (n+1)f(n+1) - nf(n) \quad (6)$$

and the general increment of order 2 is

$$I_2(n) = I_1(n+1) - I_1(n) \quad (7)$$

which is equal to, by (6),

$$I_2(n) = (n+2)f(n+2) - 2(n+1)f(n+1) + nf(n) \quad (8)$$

Some concrete examples.

(i) For the threshold index $n = t$ ($f(n) = C$, a constant, for all n) we found already that $I_1(n) = C$ and $I_2(n) = 0$ for all n.

(ii) $f(n) = n + a$, for all n, where a is constant (>0).
Here by (6),

$$\begin{aligned} I_1(n) &= (n+1)(n+a+1) - n(n+a) \\ I_1(n) &= 2n + a + 1 \end{aligned}$$

and by (7)

$$\begin{aligned} I_2(n) &= 2(n+1) + a + 1 - 2n - a - 1 \\ I_2(n) &= 2 \end{aligned} \quad (9)$$

for all n. So the h-index is not the only index with an increment of order 2 equal to 2. In fact, in the sequel it will turn out that many other functions $f(n)$ yield an $I_2(n) = 2$. Note that, for $a = 0$, we obtain the already established result (3).

(iii) $f(n) = an$, for all n, where a is constant (>0): the generalized Wu-index (Egghe, 2011).