The single publication H-index of papers in the Hirsch-core of a researcher and the indirect H-index

by

L. Egghe

Universiteit Hasselt (UHasselt), Campus Diepenbeek, Agoralaan, B-3590 Diepenbeek, Belgium[1]
and
Universiteit Antwerpen (UA), IBW, Stadscampus, Venusstraat 35, B-2000 Antwerpen, Belgium
leo.egghe@uhasselt.be

ABSTRACT

The single publication H-index of Schubert is applied to the papers in the Hirsch-core of a researcher, journal or topic. Four practical examples are given and regularities are explained: the regression line of the single publication H-index of the ranked papers in the Hirsch-core is decreasing.

We propose two measures of indirect citation impact: the average of the single publication H-indices of the papers in the Hirsch-core and the H-index of these single publication H-indices, defined as the indirect H-index. Formulae for these indirect citation impact measures are given in the Lotkaian context.

[1] Permanent address
Key words and phrases: single publication H-index, Hirsch core, indirect impact measure, indirect H-index

Introduction

The single publication H-index was introduced in Schubert (2009), for assessing single publications. Its definition is as follows. For a fixed publication, consider all publications that cite this fixed publication. If we consider this set of citing publications, one can count the citations to theses publications and rank them in decreasing order of received citations. The h-index of this ranked list (i.e. the largest rank r such that all publications receive at least r citations) is called the single publication H-index of the fixed publication.

The main idea behind this definition is that not only direct impact of a publication is important (direct citations to this publication) but also indirect impact, i.e. the citations to papers that cite this single publication.

The main topic of this paper is the following. Consider a set of papers (e.g. of a researcher, journal or topic) of which we can calculate the classical Hirsch-index (h-index) based on received citations. For this (easy) calculation, papers are ranked in decreasing order of received citations. For each paper in this ranked list we can calculate the single publication H-index as defined above. So, for each paper (say in the h-core) we have a single publication h-index, denoted $H_1,...,H_h$ (we distinguish between h, the h-index of the set of papers and H_i, the single publication H-index of the ith paper in this h-core ($i = 1,...,h$)).

Let us give a concrete example: the publications of this author and their citations on April 21st 2011 according to Thomson Reuters' Web of Science (WoS). In Table 1 we can see that the h-index of Egghe is $h = 19$.

Table 1. h-index of Egghe (on April 21st 2011) and single publication H-indices of the papers in the h-core

r	**# cit.**	**H_r**
1	172	19
2	98	13
3	85	15
4	59	10
5	57	20
6	46	15
7	40	12
8	39	13
9	33	10
10	32	7
11	27	8
12	25	8
13	22	6
14	21	7
15	20	5
16	20	7
17	20	7
18	20	10
19	19	7
20	18	

The single publication H-indices of these 19 papers is calculated as follows. In the WoS we ask for the papers of Egghe to be ranked in decreasing order of received citations. When clicking on each paper we can ask for seeing all papers that cite this paper (e.g. for the paper on rank $r=1$ we ask for retrieving all 172 citing papers). Then we ask for putting these citing papers in decreasing order of received citations from which it is easy to calculate the single publication H-index (e.g. for the first paper $H_1 = 19$, for the second paper $H_2 = 13$ and so on). This procedure is very simple and can be executed in a few minutes.

In Thor and Bornmann (2011) one presents a web application where the single publication H-index can be automatically calculated for any publication indexed in Google Scholar.

It is clear that we can expect higher H_r-values for lower r ranks. This will be studied theoretically in the next section based on the Lotkaian framework (see e.g. Egghe (2005)). Also a formula for H_h (the single publication H-index of the paper on rank $r = h$) is presented.

In section 3 we study 4 examples: the example of Egghe above, the example of Jean Bourgain (Field medalist in mathematics), the example of the journal "Scientometrics" (period 2006-2011) and the example of the relatively new journal "Journal of Informetrics" (period 2007 (first year of publication) -2011), where 2011 is limited to the dates of experimentation. Regression lines of r versus H_r confirm the decreasing relationship in all 4 cases.

In Section 4 we present two new measures of indirect citations impact: $\overline{H}$, the average of the H_i-values in the h-core and IH, the indirect H-index of the researcher, journal or topic, being the Hirsch-index of the $H_1, H_2, ..., H_h$-values. We calculate $\overline{H}$ and IH for the given examples and present formulae for $\overline{H}$ and IH in the Lotkaian framework.

The paper ends with a conclusion and open problems section.

Theoretical considerations in a Lotkaian framework

The example in Table 1 (Egghe data) can be generalized as follows. The first two columns represent a classical rank-order distribution of papers and their received citations (in decreasing order of received citations). Denote by $g(r)$ this number of received citations. In a Lotkaian framework this is the law of Zipf

$$g(r) = \frac{B}{r^\beta} \tag{1}$$

$B, \beta > 0$. This is so because of the following. Denote by $f(j)$ the number of papers with j citations. In a continuous framework we have (see Egghe (2005), Exercise II.2.2.6 or Egghe and Rousseau (2006), Appendix, where also a proof is presented).

Proposition 1:

The following assertions are equivalent

(i) $$f(j) = \frac{C}{j^\alpha} \tag{2}$$

$C > 0, \alpha > 1$ (constants) and $j \geq 1$

(ii) $$g(r)=\frac{B}{r^{\beta}} \tag{3}$$

(the law of Zipf (1) above) $B,\beta>0$ (constants) and $r\in\left]0,T\right]$ where T is the total number of papers. Moreover, the relations between the parameters are

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

$$\beta=\frac{1}{\alpha-1} \tag{5}$$

Note that it follows from the definition of $f(j)$ that, if $\alpha>1$

$$T=\int_1^{\infty} f(j)\,dj=\frac{C}{\alpha-1} \tag{6}$$

implying that (4) also reads

$$B=T^{\frac{1}{\alpha-1}} \tag{7}$$

In Egghe and Rousseau (2006) we proved that in such systems, the h-index equals

$$h=T^{1/\alpha} \tag{8}$$

For each r, the number $g(r)$ is the number of citations to the paper on rank r. In the definition of single paper h-index, we consider these $g(r)$ papers and citations to these $g(r)$ papers. Hence, for each r, we again have a classical rank-order distribution: e.g. for $r=1$ in Table 1, we have 172 papers which can be arranged in decreasing order of citations that they received. Hence, for each r we have similar functions as in (2) and (3) which we can denote

$$\varphi(j')=\frac{C'}{j'^{\alpha'}} \tag{9}$$

$C'>0,\alpha'>1, j'\geq 1$ and

$$\gamma(r')=\frac{B'}{r'^{\beta'}} \tag{10}$$

$B',\beta>0$ and $r'\in\left]0,T'_r\right[$ where here

$$T_r'=g(r) \tag{11}$$

Formula (11) is the key relation in our single publication H-index model of papers in the h-core: the citing papers (to the paper on rank r) become cited papers in the definition of single publication H-index. According to this definition and applying (8) to this case, we have that the single publication H-index of the paper on rank r equals

$$H_r = g(r)^{1/\alpha'} \tag{12}$$

Note that, strictly speaking, the new parameters α' and β' depend on the rank r. We have supposed that they are independent of r as a first approximation and for the sake of simplicity.

So we have proved the following result.

Proposition 2

In a system where papers are ranked in decreasing order of the number of received citations, for every rank r, we have that the paper on this rank has the single publication H-index H_r where

$$H_r = g(r)^{1/\alpha'} \tag{13}$$

where $g(r)$ is given by (3) and where α' is Lotka's exponent in (9) (of the Lotkaian system of the citing papers on rank r now treated as cited papers). Hence H_r is a decreasing function of r (by (3)).

Corollary 3

For all r,

$$H_r < g(r) \tag{14}$$

Proof

This follows from (11) and the fact that $\alpha' > 1$. □

Corollary 4

For $r = h$, we have

$$H_h = h^{1/\alpha'} \tag{15}$$

, an increasing function of h.

Proof

By (13), we have that

$$H_h = g(h)^{1/\alpha'} \tag{16}$$

By the very definition of the h-index we have $h = g(h)$, from which (15) follows. □

Corollary 5

For $r = h$, we have

$$H_h < h \tag{17}$$

Proof

This follows readily from (15) and the fact that $\alpha' > 1$. □

Combining (8) with (15) we also have

$$H_h = T^{\frac{1}{\alpha\alpha'}} \tag{18}$$

Although it is not true in all cases we will find, in most cases that $H_r \leq h$. In the next section we will give practical evidence for the theoretical results.

Practical examples

I The example of Egghe (Table 1)

We see that (14) is verified for all r. Note that, theoretically, H_r can have any value due to the fact that the $g(r)$ citing papers (citing the paper on rank r) can be cited by other papers, independent of the existing $g(r)$-system. But this example shows that the found theoretical regularities apply due to the (more or less) Lotkaian nature of the other systems on which the single publication H-index H_r is calculated.

It is clear that $H_h = 7 < h = 19$. Using (15) leads to an approximate α'-value

$$\alpha' = \frac{\ln h}{\ln H_h} = 1.5131 \tag{19}$$

The sequence $(H_r)_{r=1,\ldots,h}$ is not decreasing but has a decreasing regression line as Fig.1 shows. Its equation is

$$H = 16.4912 - 0.601754r \tag{20}$$

This shows that direct citations have a relation with the indirect citations as revealed through the single publication H-index. This will lead to two new impact measures of indirect citation (one of them being defined as the indirect H-index – see the next section).

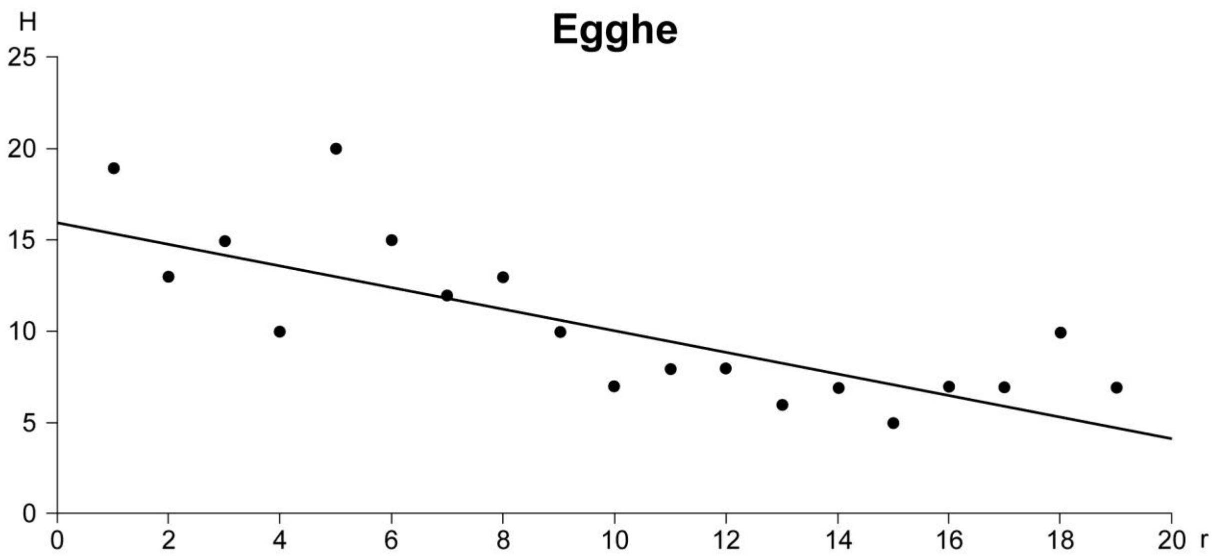

Fig.1. H_r versus r for the Egghe data

II The example of J. Bourgain (Table 2)

J. Bourgain, a Belgian mathematician, but now working in the Institute for Advanced Study in Princeton, won the Fields medal in 1994 (the Fields medals are the equivalent of the "Nobel Prices" in mathematics). We were interested if similar properties of H_r can be found in this case of extreme performance and impact – see Table 2.

Table 2. h-index of J. Bourgain (on April 21st 2011) and single publication H-indices of the papers in the h-core

r	**# cit.**	**H_r**
1	174	23
2	164	22
3	98	19
4	94	12
5	82	15
6	77	13
7	76	12
8	71	14
9	70	12
10	67	15
11	63	8
12	62	7
13	59	11
14	57	10
15	56	12
16	52	12
17	49	9
18	48	11
19	48	12
20	46	10
21	44	10
22	43	9
23	42	9
24	40	12
25	40	8
26	40	11
27	39	7
28	38	6
29	38	5
30	37	8
31	35	9
32	35	10
33	35	9
34	35	8
35	34	

We again see that (14) is verified for all r. It is clear that $h = 34$ and $H_h = 8$. Using (15) leads to an approximate α'-value

$$\alpha' = \frac{\ln h}{\ln H_h} = 1.6958 \tag{19}$$

The sequence $(H_r)_{r=1,\ldots,h}$ is not decreasing but has a decreasing regression line as Fig. 2 shows. Its equation is

$$H = 16.3422 - 0.295187r \tag{22}$$

Note finally that $H_r < h$ for all r.

So we obtain similar conclusions as in the Egghe-case showing again the relation between direct and indirect citations.

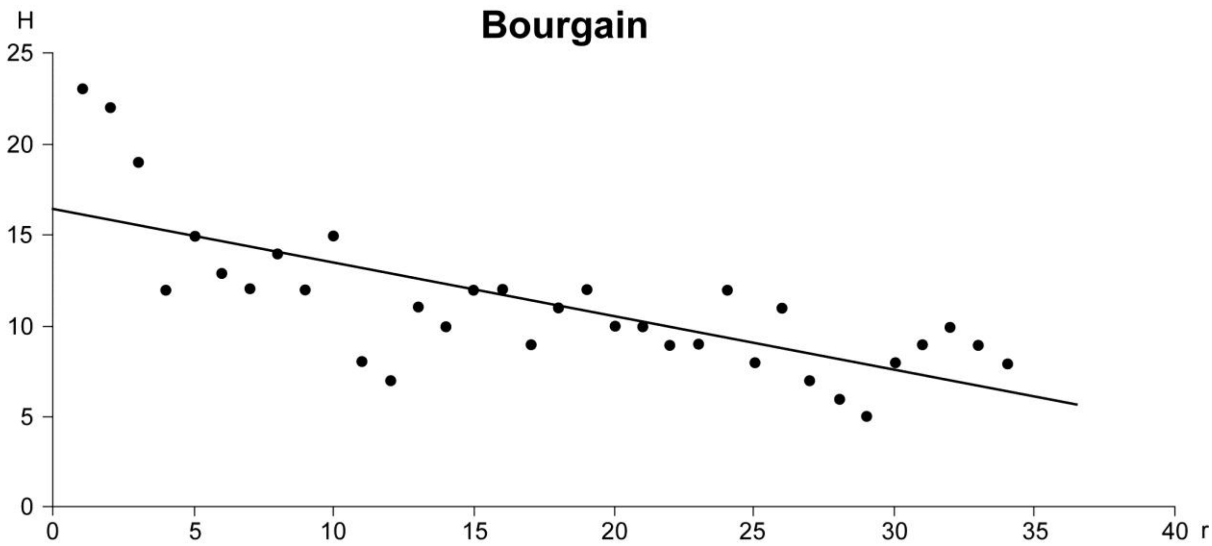

Fig.2. H_r versus r for the Bourgain data

III The example of the journal "Scientometrics" (2006-2011)

In Table 3 we present the data on H_r for the journal "Scientometrics", limited to the period 2006 – April 21, 2011 in order to limit the size of the Table.

Table 3. h-index of the journal "Scientometrics" (2006 – April 21, 2011) (on April 21st 2011) and single publication H-indices of the papers in the h-core

r	**# cit.**	**H_r**
1	172	19
2	137	19
3	90	19
4	89	14
5	85	15
6	84	20
7	82	11
8	52	6
9	51	7
10	50	13
11	45	6
12	44	11
13	43	7
14	37	11
15	32	7
16	29	8
17	28	7
18	28	7
19	27	5
20	27	6
21	26	6
22	26	7
23	25	4
24	25	4
25	24	

We again see that (14) is verified for all r. It is clear that $h = 24$ and $H_h = 4$. Using (15) leads to an approximate α'-value

$$\alpha' = \frac{\ln h}{\ln H_h} = 2.2925 \tag{21}$$

The sequence $\left(H_r\right)_{r=1,\dots,h}$ is not decreasing but has a decreasing regression line as Fig. 3 shows. Its equation is

$$H = 17.5725 - 0.60913r \tag{24}$$

Note finally that $H_r < h$ for all r. So we obtain similar conclusions as in the previous examples.

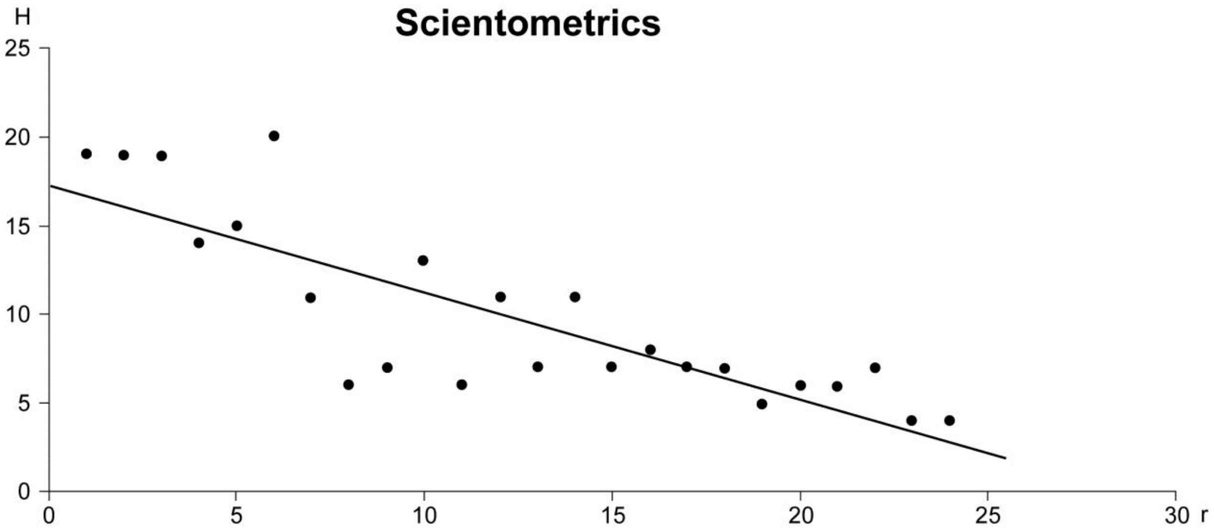

Fig.3. H_r versus r for the "Scientometrics" data

IV The case of the journal "Journal of Informetrics" (JOI)

In Table 4 we present the data on H_r for JOI.

Table 4. h-index of the journal JOI (on April 27th 2011) and single publication H-indices of the papers in the h-core

r	# cit.	H_r
1	49	6
2	39	8
3	38	7
4	37	6
5	33	9
6	31	8
7	28	5
8	25	8
9	24	8
10	24	5
11	21	3
12	19	6
13	18	4
14	18	8
15	17	4
16	14	

We again see that (14) is verified for all r. It is clear that $h = 15$ and $H_h = 4$. Using (15) leads to an approximate α'-value

$$\alpha' = \frac{\ln h}{\ln H_h} = 1.9534 \tag{25}$$

The sequence $(H_r)_{r=1,\dots,h}$ is not decreasing but has a decreasing regression line as Fig. 4 shows. Its equation is

$$H = 7.7619 - 0.178571r \tag{26}$$

Again $H_r < h$ for all r. We obtained the same conclusions as in the previous examples.

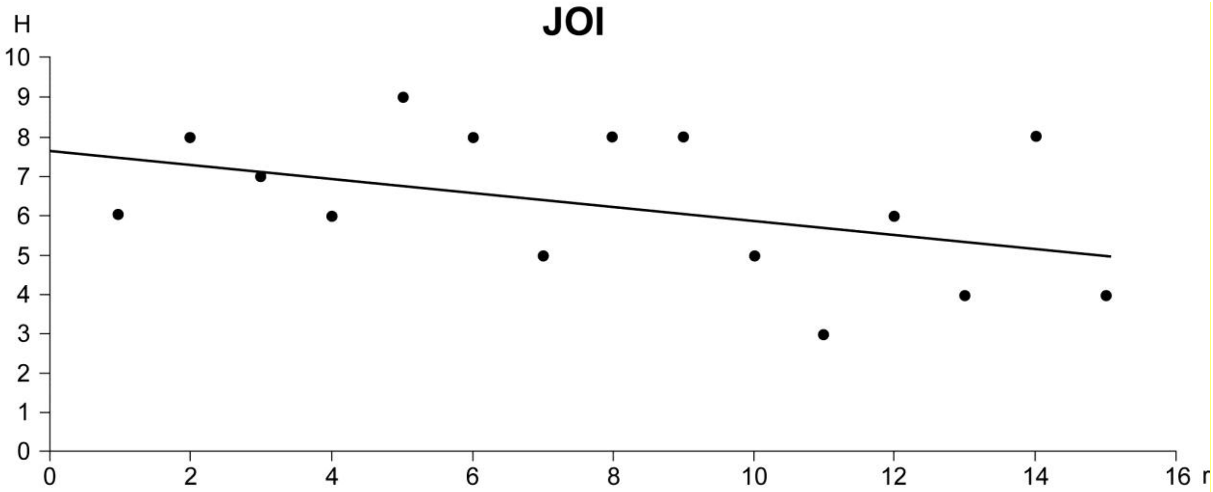

Fig.4. H_r versus r for the JOI data

The next section is devoted to the definition and study of two new measures of indirect impact of a researcher (journal, topic, …) based on the single publication H-index.

Two new measures of indirect impact

I The indirect H-index

Definition 1

The indirect H-index IH of a researcher (journal, topic) is the H-index of the values $H_1,...,H_h$.

Since the sequence $\left(H_r\right)_{r=1,...,h}$ is not necessarily decreasing, we have to put the values $H_1,...,H_h$ in decreasing order.

Table 1 gives $IH=10$ for Egghe, Table 2 gives $IH=12$ for J. Bourgain, Table 3 gives $IH=10$ for "Scientometrics" (2006-2011) and Table 4 gives $IH=7$ for JOI.

In our Lotkaian framework, we can prove the following formula for IH in function of the "direct" h-index h and the Lotkaian parameters α and α'.

Theorem 6

In the notation of the second (theoretical) section, we have

$$IH = h^{\frac{\alpha}{1+\alpha'(\alpha-1)}} \tag{27}$$

Proof

Since, by (11), for every r

$$H_r = g\left(r\right)^{1/\alpha'}$$

, the indirect H-index IH is defined by

$$g\left(IH\right)^{1/\alpha'} = IH \tag{28}$$

Using (1) this gives

$$\frac{B^{1/\alpha'}}{\left(IH\right)^{\beta/\alpha'}} = IH$$

from which we have

$$IH = B^{\frac{1}{\beta+\alpha'}} \tag{29}$$

But, by (7) and (8)

$$B = T^{\frac{1}{\alpha-1}} = T^{\frac{1}{\alpha}\frac{\alpha}{\alpha-1}} = h^{\frac{\alpha}{\alpha-1}} \tag{30}$$

So, (30) in (29) gives

$$IH = h^{\frac{\alpha}{\alpha-1}\frac{1}{\beta+\alpha'}} \tag{31}$$

But, by (5)

$$\frac{\alpha}{\alpha-1}\ \frac{1}{\beta+\alpha'} = \frac{\alpha}{\alpha-1}\ \frac{1}{\frac{1}{\alpha-1}+\alpha'} = \frac{\alpha}{1+\alpha'(\alpha-1)}$$

and this proves (25). $\square$

Corollary 7

(i) IH is an increasing function of h

(ii) $IH > H_h$

Proof

(i) follows clearly from (27) since $\alpha > 1$.

(ii) is logical from the definition of IH but also follows simply from (15) and (27). To prove that

$$IH > H_h \tag{32}$$

It suffices to prove that

$$\frac{\alpha}{1+\alpha'(\alpha-1)} > \frac{1}{\alpha'}$$

or that

$$\frac{1}{1+\alpha'(\alpha-1)} > \frac{1}{\alpha\alpha'} \tag{33}$$

But this is clear since $\alpha' > 1$. $\square$

Relation (32) can be verified in our 4 examples.

II The average H-index

Definition 2

If we take the average $\overline{H}$ of the values $H_1, \ldots, H_h$ we have the average of the single publication H-indices of the papers in the h-core.

In the discrete case we can define

$$\overline{H} = \frac{1}{h}\sum_{i=1}^{h} H_i \tag{34}$$

In our 4 examples we find $\overline{H} = 10.4737$ for Egghe, $\overline{H} = 11.1765$ for J. Bourgain, $\overline{H} = 9.9583$ for "Scientometrics" (2006-2011) and $\overline{H} = 6.3333$ for JOI.

For the continuous case, using (13), this gives

$$\overline{H} = \frac{1}{h}\int_0^h g(r)^{1/\alpha'} dr \tag{35}$$

Also for $\overline{H}$ we have a formula in function of h and the parameters α (or β) and α'.

<u>**Theorem 7**</u>

In the notation of the second (theoretical) section, we have, if $\alpha > 2$

$$\overline{H} = \frac{1}{1-\frac{\beta}{\alpha'}} h^{\frac{1}{\alpha'}} \tag{36}$$

<u>**Proof**</u>

By (35) and (3), we have

$$\overline{H} = \frac{1}{h}\int_0^h \frac{B^{1/\alpha'}}{r^{\beta/\alpha'}} dr \tag{37}$$

But, if $\alpha > 2$ we have $\beta < 1$ and since $\alpha' > 1$ we that $0 < \frac{\beta}{\alpha'} < 1$ and hence (37) yields

$$\overline{H} = \frac{B^{1/\alpha'}}{1-\frac{\beta}{\alpha'}} \frac{1}{h^{\beta/\alpha'}} \tag{38}$$

However, B depends on h and α as follows: by (4), (6) and (8) we have

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

$$= \left(\frac{C}{\alpha-1}\right)^{1/\alpha}$$

$$= \left(\frac{C}{\alpha-1}\right)^{\frac{1}{\alpha-1}\frac{\alpha-1}{\alpha}}$$

$$= B^{\frac{\alpha-1}{\alpha}} \tag{39}$$

So (39) implies

$$B^{1/\alpha'} = h^{\frac{\alpha}{(\alpha-1)\alpha'}}$$

So (38) becomes

$$\overline{H} = \frac{1}{1-\frac{\beta}{\alpha'}} h^{\frac{\alpha}{(\alpha-1)\alpha'} - \frac{\beta}{\alpha'}} \tag{40}$$

But

$$\frac{\alpha}{(\alpha-1)\alpha'} - \frac{\beta}{\alpha'} = \frac{\alpha-\beta(\alpha-1)}{\alpha'(\alpha-1)} = \frac{1}{\alpha'} \tag{41}$$

by (5). (40) and (41) yield (36). □

Corollary 8

(i) $\overline{H}$ is an increasing function of h

(ii)
$$\overline{H} = \frac{1}{1-\frac{\beta}{\alpha'}} H_h \tag{42}$$

and hence

(iii)
$$\overline{H} > H_h \tag{43}$$

Proof

(i) is evident since $0 < \frac{\beta}{\alpha'} < 1$.

Formula (42) follows from (15) and (iii) again follows from $0 < \frac{\beta}{\alpha'} < 1$. □

Using (5) we also have that

$$\overline{H} = \frac{\alpha'(\alpha-1)}{\alpha'(\alpha-1)-1} h^{1/\alpha'} \tag{44}$$

We see that all 4 examples satisfy (43).

Conclusions and suggestions for further research

The novelty of this paper is the introduction of the interesting single publication H-index (Schubert (2009)) into the framework of the papers in the h-core of a researcher (journal, topic, …).

Four practical cases are studied and we show that they share the same properties. A rationale for these properties in the Lotkaian framework is presented. The fact that these single publication H-indices decrease in the rank r shows that there is a link between direct and indirect citation impact.

Therefore we introduced the indirect H-index IH being the H-index of the values $H_1,...,H_h$. We also introduce the average H-index $\overline{H}$ being the average of the values $H_1,...,H_h$. For both indirect impact measures we present a formula in the Lotkaian framework.

We feel that both indirect impact measures IH and $\overline{H}$ are basic for the assessment of indirect impact and hence we encourage researchers to produce further practical examples of larger bodies of papers (e.g. journals, institutes, …).

Also this theory and examples could be extended to other impact measures such as the g-index (Egghe (2006)) or the R-index (Jin et al. (2007)) or other impact measures (see the review Egghe (2010) and references therein.

References

L. Egghe (2005). Power Laws in the Information Production Process: Lotkaian Informetrics. Elsevier, Oxford, UK.

L. Egghe (2006). Theory and practice of the g-index. Scientometrics 69(1), 131-152.

L. Egghe (2010). The Hirsch-index and related impact measures. Annual Review of Information Science and Technology, Volume 44 (B. Cronin, ed.), 65-114, Information Today, Medford, New Jersey, USA.

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

B. Jin, L. Liang, R. Rousseau and L. Egghe (2007). The R- and AR-indices: Complementing the h-index. Chinese Science Bulletin 52(6), 855-863.

A. Schubert (2009). Using the h-index for assessing single publications. Scientometrics 78(3), 559-565.

A. Thor and L. Bornmann (2011). The calculation of the single publication h-index and related performance measures. A web application based on Google Scholar data. Online Information Review 35(2), 291-300.