A disadvantage of h-type indices for comparing the citation impact of two researchers

Peer-reviewed author version

EGGHE, Leo (2011) A disadvantage of h-type indices for comparing the citation impact of two researchers. In: RESEARCH EVALUATION, 20(4), p. 341-346.

DOI: 10.3152/095820211X13164389670356
Handle: http://hdl.handle.net/1942/13009

A disadvantage of h-type indices for comparing the citation impact of two researchers

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

We show that the rank-frequency functions of two researchers usually intersect. As a consequence of this, different h-type indices can conclude on different impact judgements of the two researchers. Also in this paper a new indicator is proposed: the average number of citations per paper in the papers whose ranks are smaller than or equal to the intersection point of their two rank-frequency functions. The theoretical derivations are illustrated using an empirical example.

[1] Permanent address
Key words and phrases: h-type index, comparing two researchers

Introduction

Comparing the citation impact of two researchers is an important but difficult issue, e.g. in the assessment preceding a decision for a fixed position (tenure track) or for the allocation of research funds.

An example might be helpful. Two researchers both have an h-index (see below for exact definitions) that is at least 10. However, the ten best-cited papers of the first one are cited more highly than their counterparts among the ten best-cited papers of the second one, although the second researcher receives a somewhat higher number of citations for his other papers. Thus, the rank distributions of both sets of papers intersect after rank $r = 10$. A new indicator is proposed that compares the average impact of the top papers of both researchers, as indicated by the intersection point. This "truncated" average is in line with the h-index "philosophy" of using only the citation numbers of the most highly cited papers. Like the h-index, this new indicator is easy to calculate. Below, we show that the rank-frequency distributions of two researchers usually intersect (see also below for exact definitions).

The main goal of this paper is to make clear that generalized h-type indices (such as the generalized Wu- and Kosmulski-indices) do not measure the scientific impact of researchers in a uniform way. To be more concrete: when we want to compare the impact of two researchers A and B, we give explicite examples (both by theoretical and empirical examples) of such indices (dependent on a parameter) where researcher A is evaluated as having more impact than researcher B, while we have other indices where the opposite conclusion must be made.

Here we do not deal with problems as: comparing young and senior scientists, comparing scientists from different scientific fields, and so on. But even when we have scientists from the same field and anciennity, there remain problems of comparing their scientific merits.

The basic "tool" for comparing two researchers is their rank-frequency function (we can generalize this to comparing two objects such as journals or institutes, …). The rank-frequency function of a researcher is constructed as follows: rank all publications of a researcher in decreasing order of the number of citations that these publications have received (in a certain period). Let us call r the rank of the publication. Then the rank-frequency function, denoted $g(r)$, is the function that maps r onto the number of received citations of the publication on rank r.

A classical (but not so old) way of measuring the impact of a researcher, using the rank-frequency function is by applying a Hirsch-type (or h-type) index to this rank-frequency function. The Hirsch-index (or h-index) is well-known although introduced not so long ago in Hirsch (2005). A researcher has h-index h if $r = h$ is the highest rank such that all publications on ranks $1, 2, ..., h$ have at least h citations. This boils down to intersecting the rank-frequency function $g(r)$ with the first bissectrix and reading the obtained rank $r = h$.

Problems with the use of the h-index as an indicator of impact have been mentioned in several publications – see the review Egghe (2010b) but see also Egghe (2010c) (and the comments in the introductory paper of this special issue to celebrate the work of Anthony van Raan).

Generalizations of the h-index have been defined e.g. in Wu (2010) and Kosmulski (2006) and further generalized in Egghe (2010a) as follows. For $a > 0$, the generalized Wu-index is the highest rank $r = w_a$ such that all publications on ranks $1, ..., w_a$ have at least aw_a citations. For $a = 1$ we refind the h-index: $w_1 = h$. This boils down to intersecting the rank-frequency function $g(r)$ with the straight line $y = ar$, i.e. the straight line through the origin and with slope equal to a, and then reading the obtained rank $r = w_a$. Note that $h = w_1$. In Wu (2010) one used $a = 10$, a rather arbitrary value for a (that is why in Egghe (2010a) the generalized Wu-index w_a was studied, earlier introduced in van Eck and Waltman (2008)).

For $a > 0$, the generalized Kosmulski-index is the largest rank $r = h_a$ such that all publications on ranks $1, ..., h_a$ have at least $(h_a)^a$ citations. This boils down to intersecting the rank-frequency function $g(r)$ with the curve $y = r^a$ and then reading the obtained rank $r = h_a$. Note that for $a = 1$, $h = h_1$. In Kosmulski (2006) one used $a = 2$ and the above generalized Kosmulski-index was introduced in Egghe (2010a).

A common generalization of the generalized Wu- and Kosmulski-indices is by considering the intersection of the rank-frequency function $g(r)$ with the function $y = cr^d$ ($d = 1$: generalized Wu-index, $c = 1$: generalized Kosmulski-index, $c = d = 1$: h-index).

In general, any increasing curve can be used to define a new impact measure (see also Deineko and Woeginger (2009)). In the next section we will show that all these indices have disadvantages when comparing the impact of 2 researchers, hence based on the rank-frequency functions of the 2 researchers. This is done by studying the possible interrelationships of these rank-frequency functions, assuming that they satisfy Zipf's law. This is a very classical assumption since hereby we assume that we are in Lotkaian systems, Egghe (2005). That means that we assume that the two rank-frequency functions (denoted $g(r)$ and $g^*(r)$) satisfy

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

and

$$g^*(r) = \frac{B^*}{r^{\beta^*}} \tag{2}$$

where $B, B^* > 0$, $\beta, \beta^* > 0$ and where $r \geq 0$ (generalizing the discrete ranks $r = 1, 2, 3, ...$ to continuous values for ease of calculation).

We prove that, in all cases where $\beta \neq \beta^*$, curves (1) and (2) intersect. As a consequence, dependent on the value of a in the generalized Wu- and Kosmulski-index (w_a and h_a, respectively) we conclude that the first researcher has more impact than the second one for certain values of a and the opposite conclusion is also drawn (for other values of a). Even the simple h-index can conclude that the first researcher has less, more or equal impact than the second researcher, dependent on the intersection point of the curves (1) and (2).

We conclude that none of these indices are advisable to use in the comparison of two researchers. We comment on the use of the g-index in this connection and we also propose another comparison method that is not of h-type but based on the two rank-frequency functions directly: the average number of citations per paper in the papers which ranks are smaller than or equal to the intersection point of their two rank-frequency functions.

These conclusions are also valid if $g(r)$ and $g^*(r)$ are not Zipfian since these curves still can intersect.

An example of such intersecting rank-frequency curves is given by the publication-citation data of T. Braun and H. Small (data from Egghe (2006)), where we show that

(i) the generalized Wu- and Kosmulski-indices lead to contradicting conclusions with respect to the impact of the two researchers

(ii) the g-index is a better impact measure than the the h-index (but this has been remarked before in many other papers (see e.g. the review Egghe (2010b)). We acknowledge, as suggested by one of the referees, that when a parameter would be introduced in the g-index, similar problems can occur as in the case of the Wu- and Kosmulski-indices.

(iii) the newly proposed indicator (see above) is also capable of more correctly estimating the impact of two researchers (in their direct comparison).

Study of the interrelations of the curves (1) and (2) and conclusions for the use of h-type indices.

First we study when and where the curves (1) and (2) intersect : we have $g(r) = g^*(r)$ if and only if

$$\frac{B}{r^{\beta}} = \frac{B^*}{r^{\beta^*}} \tag{3}$$

This is valid for

$$r = r_0 = \left(\frac{B^*}{B}\right)^{1/\left(\beta^* - \beta\right)} \tag{4}$$

If $\beta = \beta^*$ the number r_0 does not exist: the curves $g(r)$ and $g^*(r)$ are "parallel" (in the sense that they do not intersect) and hence one of them is always strictly above the other one (unless $B = B^*$ in which case both graphs are the same). We have a situation as in Fig.1.

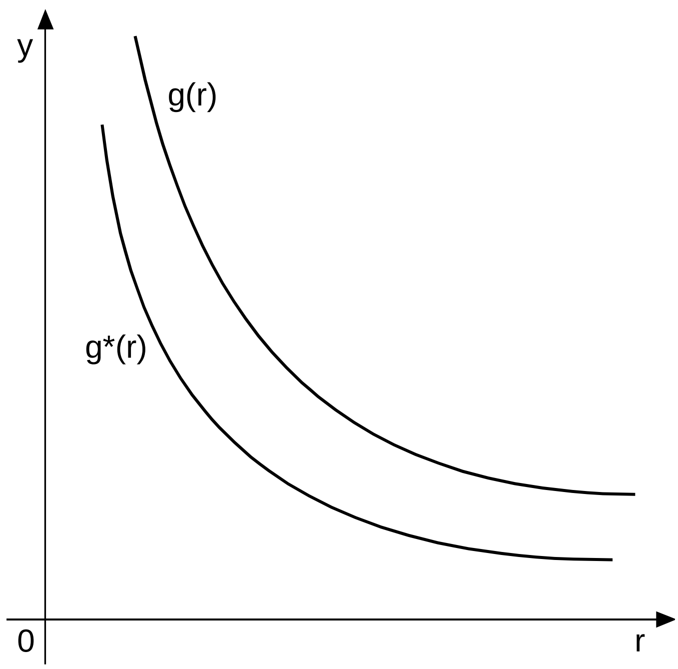

Fig.1. The case of "parallel" curves $g(r)$ and $g^*(r)$

In this case, no h-type index discussed above has a disadvantage: all intersections with $y = ar$ or $y = r^a$ yield that (say) the first researcher ($g(r)$) has more impact than the second one ($g^*(r)$).

But in most cases we will have that $\beta \neq \beta^*$, in which case the rank $r = r_0$ in (4) exists as a finite strictly positive value. Now there are two possible situations.

(a) The curves $g(r)$ and $g^*(r)$ intersect, meaning that the curve which is above the other one on the interval $]0, r_0[$ will now be below the other one on $]r_0, +\infty[$.

(b) The curves $g(r)$ and $g^*(r)$ are tangent in r_0, meaning that the curve which is above the other one on the interval $]0, r_0[$ remains above the other one on $]r_0, +\infty[$.

We will now prove that case (b) does not occur. For this we study the evolution (over $r > 0$) of the difference $g(r) - g^*(r)$. Upon interchanging $g(r)$ with $g^*(r)$ we can suppose $\beta \geq \beta^*$ and $B > B^*$ in case $\beta = \beta^*$. We have

$$\varphi(r) = g(r) - g^*(r) = \frac{B}{r^\beta} - \frac{B^*}{r^{\beta^*}} \tag{5}$$

Hence

$$\varphi'(r) = B^*\beta^* r^{-\beta^*-1} - B\beta r^{-\beta-1} \tag{6}$$

Hence $\varphi'(r) = 0$ in

$$r = r_1 = \left(\frac{B^*\beta^*}{B\beta}\right)^{1/\left(\beta^*-\beta\right)} \tag{7}$$

This point does not exist if $\beta = \beta^*$, hence in case of Fig.1.

In case $\beta \neq \beta^*$, we have $\beta^* < \beta$ since we supposed $\beta^* \leq \beta$. Now (7) reads

$r_1 = \left(\frac{B\beta}{B^*\beta^*}\right)^{1/\left(\beta-\beta^*\right)}$ from which $r_1 > r_0$ follows.

The fact that $r_1 > r_0$ excludes case (b) leaving case (a) as the only possible case. In case (b), $\varphi(r)$ decreases from $+\infty$ (this will be proved further on) until 0 in r_0 and then starts increasing again. Hence φ has its minimum in r_0, hence $r_0 = r_1$, contradicting the above finding. In case (a), φ decreases from $+\infty$ (this will be proved further on) up to 0 in $r = r_0$ and decreases further (hence $\varphi(r) < 0$ and $\varphi'(r) < 0$) but since, for $r \to +\infty$, $\varphi(r) \to 0$ necessarily (since both curves $g(r)$ and $g^*(r)$ go to 0) and since the function $\varphi'(r)$ is continuous, there must be a point $r_1 > r_0$ where $\varphi'(r) = 0$ and this point is the minimum for φ. This is exactly what we found above.

We still have to show that

$$\lim_{\substack{r\to 0\\ >}} \varphi(r) = +\infty \tag{8}$$

Indeed

$$\begin{aligned}
\lim_{\substack{r\to 0\\ >}} \varphi(r) &= \lim_{\substack{r\to 0\\ >}} \left(\frac{B}{r^{\beta}} - \frac{B^*}{r^{\beta^*}} \right) \\
&= \lim_{\substack{r\to 0\\ >}} \frac{1}{r^{\beta^*}} \left(\frac{Br^{\beta^*}}{r^{\beta}} - B^* \right) \\
&= \lim_{\substack{r\to 0\\ >}} \frac{Br^{\beta^*-\beta} - B^*}{r^{\beta^*}} \\
&= +\infty
\end{aligned}$$

since $\beta \geq \beta^*$ and $B > B^*$ if $\beta = \beta^*$

We can finally conclude that case (a) is valid, which can be depicted as in Fig.2.

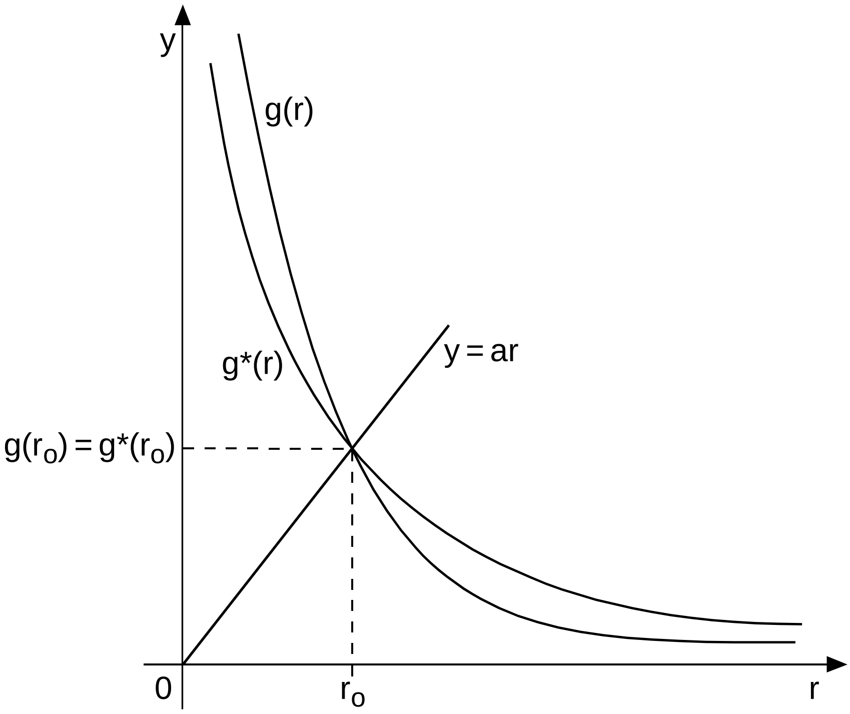

Fig.2. The case of intersecting curves $g(r)$ and $g^*(r)$

Let us calculate the value $g(r_0) = g^*(r_0)$

$$g(r_0) = \frac{B}{r_0^{\beta}}$$

$$g(r_0) = \frac{B}{\left(\left(\frac{B^*}{B}\right)^{\frac{1}{\beta^*-\beta}}\right)^{\beta}}$$

$$g(r_0) = \frac{B^{\beta^*/(\beta^*-\beta)}}{B^{*\beta/(\beta^*-\beta)}} \tag{9}$$

(and, of course, the same value is obtained when we calculate $g^*(r_0)$).

Hence the straight line connecting the origin with this intersection point has slope

$$a = \frac{g(r_0)}{r_0}$$

$$a = \frac{B}{r_0^{\beta+1}}$$

$$a = \frac{B^{(\beta^*+1)/(\beta^*-\beta)}}{B^{*(\beta+1)/(\beta^*-\beta)}} \qquad (10)$$

as is readily seen using (4) and or (9). In conclusion we see that both researchers have the same Wu-index $w_a = w_a^*$, with a as in (10).

But, as follows from Fig.2, we also have that $w_b > w_b^*$ for $b > a$ and $w_b < w_b^*$ for $b < a$ (b is a new parameter in the generalized Wu-index) which clearly shows that opposite conclusions are drawn on the 2 researchers' impact when using the generalized Wu-indices. Also, dependent on where the intersection point of both graphs is situated we can have $h < h^*$, $h = h^*$ (if $a = 1$) or $h > h^*$.

As pointed out by one of the referees, the above described disadvantage is not specific to parameterized variants of the h-index, but is relevant to any indicator that includes one or more parameters or that can be generalized to include parameters.

Although we did not draw it in Fig.2, the same negative conclusions can be drawn for the generalized Kosmulski-indices h_a, as is readily seen. In fact, the same conclusions can be drawn for any "generalized impact measure" based on the intersection of the rank-frequency function and an increasing function dependent on a parameter (e.g. the increasing functions studied in Henzinger, Suñol and Weber (2010)).

Note

All the above derivations and results are, of course, dependent on the validity of the Zipfian functions (1) and (2) (or, equivalently (see Egghe (2005), Chapter 2), Lotka's law as size-frequency function). It is known that this is a good model as a first approximation (see Egghe (2005), Chapter 1) but that small derivations exist where the rank-frequency function is not convex (as in the case of Zipf's law) but has an S-shape: a large convex part, following by a small concave part (see e.g. Egghe and Waltman (2011) and references therein).

A new indicator of impact difference between two researchers in case of intersecting rank-frequency functions.

As pointed out in the previous section, if we have two researchers for whom the general shape of Fig.1 applies, there is no difficulty of measuring the impact difference between these two researchers (in fact, any impact measure can be used). But it was also pointed out that in case of Fig.2, the measurement of the impact difference between two researchers is subject to some arbitrariness. This is true for any impact measure defined on the basis of intersection of the rank-frequency function with an increasing graph.

The g-index, Egghe (2006), is an impact measure that is not defined on the basis of an intersection of the rank-frequency function with an increasing graph. The g-index is the highest rank $r = g$ such that all papers on ranks $1,...,g$ **together**, received at least g^2 citations. As pointed out in Egghe (2006) and Egghe (2009), the g-index is capable for taking into account high numbers of citations to papers in the lowest ranks (hence with highest numbers of citations). Let us reproduce two examples given in Egghe (2006): the citation data of T. Braun and H. Small – see Table 1 and Table 2.

The notations are as follows: TC = total number of citations to the paper on rank r. The order in the tables is in decreasing order of TC, $\sum TC$ = cumulative number of citations to the first r papers, r^2 = square of the rank r. The boxes explain the calculations of the h-index and g-index.

Table 1. Citation data of T. Braun (2006)

TC	r	$\sum TC$	r^2	TC	r	$\sum TC$	r^2	TC	r	$\sum TC$	r^2
125	1	125	1	**35**	15	880	225	**23**	29	1264	841
124	2	249	4	**33**	16	913	256	**23**	30	1287	900
78	3	327	9	**32**	17	945	289	**23**	31	1310	961
66	4	393	16	**31**	18	976	324	**22**	32	1332	1024
57	5	450	25	**31**	19	1007	361	**22**	33	1354	1089
57	6	507	36	**28**	20	1035	400	**22**	34	1376	1156
55	7	562	49	**27**	21	1062	441	**22**	35	1398	1225
51	8	613	64	**27**	22	1089	484	**21**	36	1419	1296
43	9	656	81	**27**	23	1116	529	**21**	37	1440	1369
42	10	698	100	**26**	24	1142	576	**20**	38	1460	1444
38	11	736	121	**26**	25	1168	625	**20**	39	1480	1521
37	12	773	144	**25**	26	1193	676	.	.	.	.
37	13	810	169	**25**	27	1218	729	.	.	.	.
35	14	845	196	**23**	28	1241	784	.	.	.	.

Table 2. Citation data of H. Small (2006)

TC	r	$\sum TC$	r^2	TC	r	$\sum TC$	r^2	TC	r	$\sum TC$	r^2
305	1	305	1	**25**	15	1371	225	**5**	29	1536	841
239	2	544	4	**22**	16	1393	256	**5**	30	1541	900
127	3	671	9	**22**	17	1415	289	**3**	31	1544	961
109	4	780	16	**18**	18	1433	324	**3**	32	1547	1024
86	5	866	25	**18**	19	1451	361	**2**	33	1549	1089
80	6	946	36	**15**	20	1466	400	**2**	34	1551	1156
77	7	1023	49	**12**	21	1478	441	**2**	35	1553	1225
75	8	1098	64	**10**	22	1488	484	**1**	36	1554	1296
67	9	1165	81	**9**	23	1497	529	**1**	37	1555	1369
49	10	1214	100	**8**	24	1505	576	**1**	38	1556	1444
44	11	1258	121	**8**	25	1513	625	**1**	39	1557	1521
36	12	1294	144	**7**	26	1520	676	**1**	40	1558	1600
26	13	1320	169	**6**	27	1526	729	.	.	.	.
26	14	1346	196	**5**	28	1531	784	.	.	.	.
								.	.	.	.

We see that, in 2006, T. Braun had an h-index of 26 while H. Small had an h-index of 18. Nevertheless we see that , up to rank $r = 11$, H. Small received (much) more citations than T. Braun but T. Braun continuous to have higher number of citations to papers of higher ranks. From the Tables 1 and 2, we see that T. Braun had a g-index of 38, almost the same as H.

Small who had a g-index of 39. So, here, H. Small was "compensated" for the higher number of citations to papers of the 11 ranks.

This is a typical case of intersecting rank-frequency functions (as in Fig.2). The (continuous) intersection point is between ranks 11 and 12. An interesting indicator could be: the average number of citations per paper in the papers whose ranks are smaller than or equal to the intersection point. Hence, in our example, we use the first 11 papers, yielding an average of $\frac{736}{11} = 66.9$ citations per paper for T. Braun and $\frac{1,258}{11} = 114.4$ citations for H. Small, expressing clearly the higher concentration of citations to papers of H. Small than to papers of T. Braun. This is not reflected in the overall average number of citations (which is difficult to calculate). Over the first 39 papers (which data are available in Tables 1 and 2) we have an average number of citations per paper of $\frac{1,480}{39} = 37.9$ for T. Braun and $\frac{1,557}{39} = 39.9$ for H. Small (hence close together).

Note that this indicator is easy to calculate (as opposed to the overall average number of citations per paper). This "truncated" average is in line with the h-index "philosophy" of using only the citation numbers to the highest cited papers.

We think this new indicator is worth adding to the informetric toolbox in case we have an "incomparable" situation as in Fig.2. Such an indicator, clearly, is not needed in the perfectly comparable situation as in Fig.1.

We agree with one of the referees that, according to this new indicator, the induced rankings are the same as when we only consider the highest-cited paper. But this referee also agrees with this author that this indicator yields different values than any existing indicator (and actual values are always finer than ranks deduced from them!).

<u>Note</u>: In the case of 1 researcher but considered at two different time periods, we always have a situation as in Fig.1, since citations do not disappear in time.

Conclusions and suggestions for further research

The main goal of this paper is to make clear that generalized h-type indices (such as the generalized Wu- and Kosmulski-indices) do not measure the scientific impact of researchers in a uniform way. To be more concrete: when we want to compare the impact of two researchers A and B, we give explicite examples (both by theory and empirical examples) of such indices (dependent on a parameter) where researcher A is evaluated as having more impact than researcher B, while we have other indices where the opposite conclusion must be made.

As recognized by one of the referees, this is "normal" in any system where indicators, dependent on a parameter, are used (so not only in the connexion of h-type indices) but it remains, nevertheless, important to give exact results on such ambiguities in order to show exactly where the problems are and how big these problems are. The main basic tool in studying researchers' impact is by considering their rank-frequency functions such as the ones in (1) and (2).

We have noted that, mathematically, curves (1) and (2) (the classical Zipf versions of two rank-frequency functions) usually intersect and that in this case it is not really possible (for comparing the impact of two researchers) to use h-type indices that are defined based on the intersection of the rank-frequency function and on an increasing function (such as $y = ar$ for the generalized Wu-indices and $y = r^a$ for the generalized Kosmulski-indices, where $a > 0$, including the h-index $a = 1$).

In these "incomparable" cases we suggest a new indicator: the average number of citations per paper to the papers with ranks smaller than or equal to this intersection point. It is a comparative measure of impact, expressed by the intersection point of the two rank-frequency functions. It is easy to calculate (as opposed to the overall average number of citations per paper, which is an overall measure of citation impact).

It is clear that the problem of comparing the impact of more than two researchers is a "multiple" of the problem of comparing the impact of two researchers. Although the above defined new indicator can be used, we leave open the problem of the overall ranking of these researchers according to their calculated impact.

References

Deineko, V.G. and Woeginger, G.J. 2009. A new family of scientific impact measures: The generalized Kosmulski-indices. *Scientometrics* **80**(3), 819-826.

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

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

Egghe, L. 2009. An econometric property of the g-index. *Information Processing and Management* **45**(4), 484-489.

Egghe, L. 2010a. Characterizations of the generalized Wu- and Kosmulski-indices in Lotkaian systems. Preprint.

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

Egghe, L. 2010c. Mathematical results on the H-index and H-sequence of Randić. *Research Evaluation* **19**(3), 203-207.

Egghe, L. and Waltman L. 2011. Relations between the shape of a size-frequency distribution and the shape of a rank-frequency distribution. *Information Processing and Management*, **47**(2), 238-245.

Henzinger, M., Suñol, J. and Weber I. 2010. The stability of the h-index. *Scientometrics* **84**(2) 465-479.

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**(46), 16569-16572.

Kosmulski, M. 2006. A new Hirsch-type index saves time and works equally well as the original h-index. *ISSI Newsletter* **2**(3), 4-6.

van Eck, N.J. and Waltman, L. 2008. Generalizing the h- and g-indices. *Journal of Informetrics* **2**(4), 263-271.

Wu, Q. 2010. The w-index: A measure to assess scientific impact by focusing on widely cited papers. *Journal of the American Society for Information Science and Technology* **61**(3), 609-614.