The Influence of Transformations on the h-Index and the g-Index

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

In a previous article, we introduced a general transformation on sources and one on items in an arbitrary information production process (IPP). In this article, we investigate the influence of these transformations on the h-index and on the g-index. General formulae that describe this influence are presented. These are applied to the case that the size-frequency function is Lotkaian (i.e., is a decreasing power function). We further show that the h-index of the transformed IPP belongs to the interval bounded by the two transformations of the h-index of the original IPP, and we also show that this property is not true for the g-index.

Introduction

We suppose that we have a general information production process (IPP) with $f:[a, \rho_m] \to \mathbb{R}^+$, $j \to f(j)$ as size-frequency function and with $F:[0, T] \to \mathbb{R}^+$, $r \to F(r)$ as rank-frequency function. Here j denote item densities and r rank densities; a and ρ_m are the minimum and maximum item densities and T denotes the total number of sources. We will limit ourselves to $a = 1$. Note that F usually is denoted by g (see Egghe (2005)); but, to avoid confusion with the g-index, we denote the rank-frequency function by F.

Such a general IPP can be transformed in many ways, hereby describing a possible evolution of this IPP into another one, which we will denote by using asterisks $*$: size-frequency function $f^*: [a^*, \rho_m^*] \to \mathbb{R}^+$, $j^* \to f^*(j^*)$ and rank-frequency function $F^*:[0, T^*] \to \mathbb{R}^+$, $r^* \to F^*(r^*)$.

A very general way of describing the evolution from one IPP to another one is by applying two transformations: one on the sources:

$$\begin{aligned} \psi&:[0, T] \to [0, T^*] \\ r &\to r^* = \psi(r) \end{aligned} \tag{1}$$

(ψ differentiable and increasing $\psi(0) = 0$ and $\psi(T) = T^*$) and one on the items:

$$\begin{aligned} \varphi&:[a, \rho_m] \to [a^*, \rho_m^*] \\ j &\to j^* = \varphi(j) \end{aligned} \tag{2}$$

(φ differentiable and increasing $\varphi(a) = a^*$, $\varphi(\rho_m) = \rho_m^*$). These two functions φ and ψ define the new rank-frequency function F^*:

$$F^*(r^*) = F^*(\psi(r)) = \varphi(F(r)) \tag{3}$$

for $r \in [0, T]$.

General properties are studied in Egghe (2007a) where one also presents a formula for the transformed size-frequency function f^*. We proved the following result:

$$f^*(j^*) = f(j)\frac{\psi'(F^{-1}(j))}{\varphi'(j)} \tag{4}$$

for $j^* = \varphi(j)$ (we assume φ strictly increasing so that $\varphi'(j) \neq 0$). In the case that $\psi(r) = Ar^b$ and $\varphi(j) = Bj^c$ ($A, B > 0$, $b, c > 0$ and in case f is the function of Lotka:

$$f(j) = \frac{C}{j^\alpha} \tag{5}$$

with $\rho_m = \infty$ we prove in Egghe (2007a) that

$$f^*(j^*) = \frac{D}{j^{*\delta}} \tag{6}$$

where

$$D = \frac{CAbE^{(b-1)(\alpha-1)}B^{\frac{(\alpha-1)b}{c}}}{c} \tag{7}$$

and where

$$\delta = \frac{c + (\alpha - 1)b}{c} \tag{8}$$

Received February 7, 2007; revised May 8, 2007; accepted December 19, 2007

[1]Permanent address.

• Published online 7 April 2008 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.20823

and where E is the parameter in the rank-frequency function

$$F(r) = \frac{E}{r^{\beta}} \tag{9}$$

which is equivalent with Equation 5 as is well-known; here

$$E = \left(\frac{C}{\alpha - 1}\right)^{\frac{1}{\alpha-1}} \tag{10}$$

and

$$\beta = \frac{1}{\alpha - 1} \tag{11}$$

see Egghe (2005), Exercise II.2.2.6 or Egghe and Rousseau (2006a) where the complete proof is given.

The importance of these basic transformation results has been described in two articles. In Egghe and Rousseau (2006b), one shows that IPPs which grow such that the number of sources remains constant but where the number of items in each source grows extensively, have low Lotka exponents α. This proof is given using Equation 8 and confirmed in several examples of communities such as country sizes or municipality sizes or even database sizes. Cothey (2007) uses the same Equation 8 to predict certain evolutions in (parts of) WWW.

We can now wonder what will be the effect of the above transformations on the h-index and g-index of an IPP. We first recall the definitions of these indexes. Hirsch (2005) defines the h-index as the largest rank h of a source such that this source (and hence also the sources on a lower rank) has h or more items (Hirsch uses the paper-citation terminology hereby defining the h-index for an author). Since here we work in the continuous case in which we only have continuous functions (hence where all values between two range values are attained), we have the following easy definition of the h-index:

$$F(h) = h \tag{12}$$

In Egghe and Rousseau (2006a), we show that h always uniquely exists. An alternative definition is

$$\int_h^{\infty} f(j)dj = h \tag{13}$$

as is readily seen.

The h-index, introduced only 2 years ago, has become a real hype in and even outside informetrics: Ball (2005, 2007), Bornmann and Daniel (2005, 2007a), Braun, Glänzel, and Schubert (2005, 2006; introducing the h-index for journals, yielding a new journal indicator to be preferred above the impact factor—see Miller (2006), Glänzel (2006a,b), van Raan (2006), Bar-Ilan (2006), Rousseau (2007a), Burrell (2007a,b), Glänzel and Persson (2005), Egghe (2007b), Saad (2006), Oppenheim (2007), Hirsch (2007), Barendse (2007), Wan, Hua and Rousseau (2007), Rao and Rousseau (2007), Vinkler (2007), Vanclay (2007) and see also the articles in the special issue on the Hirsch index in Journal of Informetrics 1(3), 2007: Schubert and Glänzel (2007), Beirlant, Glänzel, Carbonez, and Leemans (2007), Costas and Bordons (2007), and Bornmann and Daniel (2007b).

Banks (2006) introduces the interesting notion of the h-index for topics and compounds—see also Egghe and Rao (2007a) and the STIMULATE6 Group (2007). Schubert (2007) and Prathap (2006) introduce h-indices for institutions via the notion of "successive h-indices" (see also Egghe, 2007c and Egghe and Rao, 2007b). Let us finally note that both the Web of Science and Scopus offer the h-index in their databases (remarkably quick after its introduction in 2005!).

According to Egghe (2006b; see also Egghe, 2006a,c), the h-index has (at least) one disadvantage: It does not take into account the exact number of citations of the first h papers. In other words, once an article belongs to the h most cited papers, it does not matter anymore how many citations it actually received, even when this number can be very high. Therefore, Egghe defined the g-index as the highest rank g of a source such that all the sources on this and lower ranks have together, at least g^2 items. In the continuous model, this gives the following defining equation for the g-index:

$$G(g) = \int_0^g F(r)dr = g^2 \tag{14}$$

Alternatively, g is defined via (the less simple) equations

$$\int_j^{\infty} f(j')dj' = g \tag{15}$$

$$\int_j^{\infty} j'f(j')dj' = g^2 \tag{16}$$

In Egghe (2006b), it is shown that also g uniquely exists.

Of course we assume here that $I \leq T^2$ (I = total number of items, T = total number of sources). This is not always true since, if $\rho_m = \infty$ (what we suppose in this paper), and if f is Lotkaian (Equation 5) with $\alpha > 2$, then it follows from

$$I = \int_1^{\infty} jf(j)dj$$

and

$$T = \int_1^{\infty} f(j)dj$$

that (see also Proposition II.2.1.1.1 in Egghe (2005)

DOI: 10.1002/asi

$$I = \frac{\alpha - 1}{\alpha - 2} T$$

(I is called A in Egghe (2005)), so I can be larger than T^2 if $\alpha > 2$ is close enough to 2. In practice, one can add fictitious sources with 0 items so that g can be defined beyond T. But in this theoretical article, this is no problem: We just assume that $I < T^2$ which implies

$$\frac{\alpha - 1}{\alpha - 2} < T$$

hence

$$\alpha > \frac{2T - 1}{T - 1}$$

which is almost always true if $\alpha > 2$ since T usually is large.

It is further trivial that $g \geq h$ in each IPP.

In Egghe (2006b,c), it is shown that the g-index has more discriminatory power than the h-index. This finding was also confirmed in Schreiber (2007) and Tol (2007). The h- and g-index (and some other indices) can be calculated using the software program "Publish or Perish" (see http://www.harzing.com/pop.htm).

The above explanations of the importance of transformations in IPPs and of the importance of the h- and the g-index should make it clear that the study of the influence of transformations on the h- and g-index is equally important.

In the next section the analogue definitions of the h- and g-index in the transformed system will be studied and calculated in general systems and in case we have a Lotkaian system (Equation 5).

In the third section we prove, denoting by h the h-index in the original system and by h* the h-index in the transformed system, that always

$$\varphi(h) \leq h^* \leq \psi(h) \tag{17}$$

or

$$\psi(h) \leq h^* \leq \varphi(h) \tag{18}$$

(with strict inequalities if $\psi(h) \neq \varphi(h)$). We also show by example that none of the above double inequalities are true for the g-index.

Finally we present conclusions and suggestions for further research.

General Equations for the h- and g-Index in General Transformed IPPs and in Lotkaian Systems

In the sequel, we will denote by h and g the h-index and g-index in the original IPP and by h* and g* the h-index and g-index in the transformed IPP, the transformations being given by Equations 1, 2, and 3.

*General Equations for h**

Theorem 2.1: The transformation formula for h is

$$\int_{j'=\varphi^{-1}(h^*)}^{\infty} f(j')\psi'(F^{-1}(j'))dj' = h^* \tag{19}$$

For the proof, see Appendix A.

Let us now illustrate how these basic equations can be used in the concrete calculation of h^*.

Calculation of h Given That the Original IPP is Lotkaian*

Theorem 2.2: We suppose that f is as in Equation 5, the law of Lotka. We will again suppose that the transformations are increasing power laws, an important case: for $A, B > 0$, $b, c > 0$:

$$j^* = \varphi(j) = Bj^c \tag{20}$$

$$r^* = \psi(r) = Ar^b \tag{21}$$

Then we have that

$$h^* = B^{\frac{\delta-1}{\delta}} T^{*\frac{1}{\delta}} \tag{22}$$

with δ as in (8).

For the proof, we refer to Appendix B.

This result generalizes the result

$$h = T^{\frac{1}{\alpha}} \tag{23}$$

obtained in Egghe and Rousseau (2006a).

We now turn our attention to the calculation of the g-index.

*General Equation for g**

For the sake of simplicity, we will only use Equation 14. For the transformed system, this gives

$$\int_0^{g^*} F^*(r^*)dr^* = g^{*2} \tag{24}$$

By Equations 1 and 3, we have

$$\int_0^{\psi^{-1}(g^*)} \varphi(F(r))\psi'(r)dr = g^{*2} \tag{25}$$

which is the basic general equation for g^*.

Again, we will illustrate its use, given Equations 5, 20, and 21.

Calculation of g^* Given that the Original IPP is Lotkaian

Theorem 2.3: We suppose that f is as in Equation 5 and φ and ψ are as in Equations 20 and 21. Then

DOI: 10.1002/asi

$$g^* = B^{\frac{\delta-1}{\delta}} T^{*\frac{1}{\delta}} \left(\frac{\delta - 1}{\delta - 2}\right)^{\frac{\delta-1}{\delta}} \tag{26}$$

with δ as in Equation 8.

For the proof, we refer to Appendix C.

Note that the requirement $\beta c + 1 - b < 1$ is needed for the convergence of the integral. For the same reason one must require, in the original IPP, that $\alpha > 2$ in order to be able to calculate the g-index g. If we work in systems with bounded densities, then Equation 9 is replaced by the function of Mandelbrot (see Egghe (2005) and the restrictions can be dropped. We do not follow this approach since calculations become very intricate. Note also, as is readily seen, that restriction $\beta c + 1 - b < 1$ is equivalent with $\delta > 2$ (so exactly the same requirement as $\alpha > 2$ in the original system).

Equation 26 could also have been obtained from Equation 22 together with the result on the g-index, proved in Egghe (2006b; in our notation):

$$g^* = h^* \left(\frac{\delta - 1}{\delta - 2}\right)^{\frac{\delta-1}{\delta}}$$

Qualitative Study of h^* and g^* in Comparison With h and g

Since h* and g* are the h-index and g-index of the transformed system, it would be logical that one can prove relations with $\varphi(h)$ and $\psi(h)$, respectively $\varphi(g)$ and $\psi(g)$, the transformed values of h and g, respectively the h-index and g-index of the original system. Note that h and g, by definition, can be considered as arguments of φ and ψ. This is clear from Equation 12 for h and from Equation 15 for g (implying $g \leq T$, assuming $I \leq T^2$ as we do in this article) and since the argument of φ is unbounded (since we assume $\rho_m = \infty$ in this article).

We will now show that h* is limited by $\varphi(h)$ and $\psi(h)$ but that this is not the case for g* in relation with $\varphi(g)$ and $\psi(g)$.

We first prove a Lemma.

Lemma 3.1: For all IPPs, we have

$$F^*(\psi(h)) = \varphi(F(h)) = \varphi(h) \tag{27}$$

Proof: This follows readily from (3) and by the fact that the h-index satisfies $F(h) = h$, by (12). □

Theorem 3.2: For all IPPs, we have

$$h^* = \varphi(h) = \psi(h) \tag{28}$$

or

$$\varphi(h) < h^* < \psi(h) \tag{29}$$

or

$$\psi(h) < h^* < \varphi(h) \tag{30}$$

For the proof, we refer to Appendix D.

Corollary 3.3: For all IPPs, we have

$$h = \varphi^{-1}(h^*) = \psi^{-1}(h^*) \tag{31}$$

or

$$\psi^{-1}(h^*) < h < \varphi^{-1}(h^*) \tag{32}$$

or

$$\varphi^{-1}(h^*) < h < \psi^{-1}(h^*) \tag{33}$$

Proof: Equation 31 follows from Equation 28, Equation 32 from Equation 29 and Equation 33 from Equation 30 using that φ and ψ are increasing functions. □

We will now show by example that Theorem 3.2 (hence also Corollary 3.3) is *not* true for the g-index.

Example 3.4: Take $\psi = \mathrm{Id}$, the identity function (hence $A = b = 1$ and $T = T^*$) and let $\varphi(j) = j^2$ (hence $B = 1$, $c = 2$). Let the original IPP be Lotkaian with $\alpha = 3.1$ (and $\rho_m = \infty$). We have

$$h = T^{\frac{1}{\alpha}} = T^{\frac{1}{3.1}}, g = \left(\frac{\alpha-1}{\alpha-2}\right)^{\frac{\alpha-1}{\alpha}} h = \left(\frac{2.1}{1.1}\right)^{\frac{2.1}{3.1}} T^{\frac{1}{3.1}} = 1.5496634$$

$$T^{\frac{1}{3.1}} = \psi(g), \varphi(g) = g^2 = \left(\frac{2.1}{1.1}\right)^{2\frac{2.1}{3.1}} T^{\frac{2}{3.1}} = 2.4014567 T^{\frac{2}{3.1}}.$$

By (8) we have

$$\delta = \frac{\alpha + 1}{2} \text{ and } \left(\frac{\delta - 1}{\delta - 2}\right)^{\frac{\delta-1}{\delta}} = \left(\frac{\alpha - 1}{\alpha - 3}\right)^{\frac{\alpha-1}{\alpha+1}}$$

$$\text{so } g^* = \left(\frac{\alpha - 1}{\alpha - 3}\right)^{\frac{\alpha-1}{\alpha+1}} T^{\frac{2}{\alpha+1}}$$

(by (26) and the fact that $T = T^*$ and $B = 1$). Hence

$$g^* = \left(\frac{2.1}{0.1}\right)^{\frac{2.1}{4.1}} T^{\frac{2}{4.1}} = 4.7559171 T^{\frac{2}{4.1}}.$$

It is now clear that $\psi(g) < \varphi(g)$ and $g^* > \varphi(g)$ if $4.7559171 T^{\frac{2}{4.1}} > 2.4014567 T^{\frac{2}{3.1}}$ which is true for $T < 76.896529$. So in all these cases

$$g^* \notin]\psi(g), \varphi(g)[$$

proving that Theorem 3.2 is not true for the g-index. By interchanging φ and ψ we also see that the other inequalities in Theorem 3.2 are also not true for the g-index. Note that in the above example

$$h^* \in]\psi(h), \varphi(h)[$$

DOI: 10.1002/asi

since

$$T^{\frac{2}{4.1}} \in \left]T^{\frac{1}{3.1}}, T^{\frac{2}{3.1}}\right[$$

in fact, for all α

$$T^{\frac{2}{\alpha+1}} \in \left]T^{\frac{1}{\alpha}}, T^{\frac{2}{\alpha}}\right[$$

since $\alpha > 1$, illustrating Theorem 3.2. So we see that the factor

$$\left(\frac{\delta - 1}{\delta - 2}\right)^{\frac{\delta-1}{\delta}}$$

is responsible for allowing (in some cases) g* not to belong to $]\psi(g), \varphi(g)[$. One can readily verify that, if $\alpha = 10$ in the above example we now have that $g^* \in]\psi(g), \varphi(g)[$ showing that this case can happen too. Indeed, for ψ and φ as above we have

$$g = \left(\frac{9}{8}\right)^{0.9} T^{0.1}$$

$$\varphi(g) = g^2 = \left(\frac{9}{8}\right)^{1.8} T^{0.2} = 1.2361596T^{0.2}$$

$$\psi(g) = g = 1.1118271T^{0.1}$$

$$g^* = \left(\frac{9}{7}\right)^{\frac{9}{11}} T^{\frac{2}{11}} = 1.2282875T^{\frac{2}{11}}$$

which obviously shows that $g^* \in]\psi(g), \varphi(g)[$.

We can also give an example where

$$g^* \neq \varphi(g) = \psi(g)$$

showing that also Equation 28 in Theorem 3.2 is false for the g-index. Indeed, take $\varphi(j) = j^2$, $\psi(r) = r^2$ for all j, r. We know that in this case $\delta = \alpha$ (since $b = c = 2$ and see Equation 8). Hence

$$g = \left(\frac{\alpha - 1}{\alpha - 2}\right)^{\frac{\alpha-1}{\alpha}} T^{\frac{1}{\alpha}}$$

$$g^* = \left(\frac{\alpha - 1}{\alpha - 2}\right)^{\frac{\alpha-1}{\alpha}} T^{*\frac{1}{\alpha}}$$

But $T^* = \psi(T) = T^2$, so

$$g^* = \left(\frac{\alpha - 1}{\alpha - 2}\right)^{\frac{\alpha-1}{\alpha}} T^{\frac{2}{\alpha}}$$

But

$$\varphi(g) = \psi(g) = \left(\frac{\alpha - 1}{\alpha - 2}\right)^{\frac{2(\alpha-1)}{\alpha}} T^{\frac{2}{\alpha}} \neq g^*.$$

Conclusions and Suggestions for Further Research

We considered a very general double transformation on an IPP: $j \rightarrow \varphi(j)$ for the items and $r \rightarrow \psi(r)$ for the sources such that the rank-frequency function F* of the transformed IPP relates to the rank-frequency function F of the original IPP as follows

$$F^*(r^*) = F^*(\psi(r)) = \varphi(F(r))$$

for $r \in [0, T]$.

Based on this we show that the basic equation for the h-index h* of the transformed IPP is

$$\int_{j'=\varphi^{-1}(h^*)}^{\infty} f(j')\psi'(F^{-1}(j'))dj' = h^*$$

or, equivalently,

$$\varphi^{-1}(h^*) = F(\psi^{-1}(h^*))$$

For the g-index g* of the transformed system we have the basic equation

$$\int_0^{\psi^{-1}(g^*)} \varphi(F(r))\psi'(r)dr = g^{*2}$$

These equations are then used to prove that

$$h^* = B^{\frac{\delta-1}{\delta}} T^{*\frac{1}{\delta}}$$

if we have a Lotkaian system

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

and where $\psi(r) = Ar^b$ and $\varphi(j) = Bj^c$ and where

$$\delta = \frac{c + (\alpha - 1)b}{c}$$

For g* we proved

$$g^* = \left(\frac{\delta - 1}{\delta - 2}\right)^{\frac{\delta-1}{\delta}} B^{\frac{\delta-1}{\delta}} T^{\frac{1}{\delta}}$$

for $\delta > 2$.

Finally we prove that

$$h^* = \varphi(h) = \psi(h)$$

or

$$\varphi(\mathrm{h}) < \mathrm{h}^* < \psi(\mathrm{h})$$

or

$$\psi(\mathrm{h}) < \mathrm{h}^* < \varphi(\mathrm{h})$$

for any IPP and we show that none of these inequalities or equalities are generally true for the g-index.

The transformations φ and ψ are generalizations of positive reinforcement of IPPs which belongs to linear three-dimensional informetrics theory (cf. Egghe, 2005, chap. 3; Egghe, 2004; Rousseau, 1992). We leave open to study linear three-dimensional informetrics theory from the point of view of the h- and g-index. More in particular it would be interesting to see if conclusions around the h-index and the g-index of the composed IPP can be drawn based on the h- and g-indexes of the composing IPPs.

Possibly also other transformations of IPPs can be studied and it would then be interesting to study the transformed h- and g-indexes based on the h- and g-indexes of the original IPP, as we studied here.

One referee asked the question if these results are true in the discrete setting (after all, h- and g-indices are calculated from discrete tables). Certainly the transformation formulae cannot be proved in the discrete setting. The reason why we work in the continuous setting is the calculability of the formulae. To give a simple example: Equation 23:

$$\mathrm{h} = \mathrm{T}^{\frac{1}{\alpha}}$$

was proved first in Glänzel (2006b) in an approximative way, but the continuous derivation, given in Egghe and Rousseau (2006a), is exact and more elegant.

It remains an interesting problem to prove (or disprove) the validity of one of the Equations 28, 29, or 30 in the discrete setting. To be honest, I do not know if the same result is valid in this discrete case, and I have to leave it as an open problem.

References

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

Ball, P. (2007). Achievement index climbs the ranks. Nature, 448, 16 August 2007, 737.

Banks, M.G. (2006). An extension of the Hirsch index: indexing scientific topics and compounds. Scientometrics 69(1), 161–168.

Bar-Ilan, J. (2006). H-index for Price medallists revisited. ISSI Newsletter 2(1), 3–5.

Barendse, W. (2007). The strike rate index: a new index for journal quality based on journal size and the h-index of citations. Biomedical Digital Libraries 4(3), 2007. http://www.bio-diglib.com/content/4/1/3

Beirlant, J., Glänzel, W., Carbonez, A. & Leemans, H. (2007). Scoring research output using statistical quantile plotting. Journal of Informetrics, 1(3), 185–192.

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

Bornmann, L., & Daniel, H.-D. (2007a). What do we know about the h-index? Journal of the American Society for Information Science and Technology, 58(9), 1381–1385.

Bornmann, L., & Daniel, H.-D. (2007b). Convergent validation of peer review decisions using the h-index. Extent of and reasons for type I and type II errors. Journal of Informetrics, 1(3), 204–213.

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

Braun, T., Glänzel, W., & Schubert, A. (2006). A Hirsch-type index for journals. Scientometrics, 69(1), 169–173.

Burrell, Q.L. (2007a). Hirsch's h-index: a stochastic model. Journal of Informetrics, 1(1), 16–25.

Burrell, Q.L. (2007b). On the h-index, the size of the Hirsch core and Jin's A-index. Journal of Informetrics, 1(2), 170–177.

Costas, R., & Bordons, M.(2007). The h-index: advantages, limitations and its relation with other bibliometric indicators at the micro level. Journal of Informetrics, 1(3), 193–203.

Cothey, V. (2007). Applying Egghe's general theory of the evolution of Information Production Processes to the World Wide Web. Proceedings of ISSI 2007. 11th International Conference of the International Society for Scientometrics and Informetrics, CSIC, Madrid (Spain) (D. Torres-Salinas and H.F. Moed, eds.), 231–240, 2007.

Egghe, L. (2004). Positive reinforcement and 3-dimensional informetrics. Scientometrics, 60(3), 497–509. Correction: Scientometrics, 61(2), 283.

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

Egghe, L. (2006a). How to improve the h-index. Letter. The Scientist, 20(3), 14, March.

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

Egghe, L. (2006c). An improvement of the h-index: the g-index. ISSI Newsletter 2(1), 8–9.

Egghe, L. (2007a). General evolutionary theory of information production processes and applications to the evolution of networks. Journal of Informetrics 1(2), 115–122.

Egghe, L. (2007b). Dynamic h-index: the Hirsch index in function of time. Journal of the American Society for Information Science and Technology, 58(3), 452–454.

Egghe, L. (2007c). Modelling successive h-indices. Scientometrics, to appear.

Egghe, L., & Ravichandra Rao, I.K. (2007a). The influence of the broadness of a query of a topic on its h-index: models and examples of the h-index of N-grams. Journal of the American Society for Information Science and Technology.

Egghe, L., & Ravichandra Rao, I.K. (2007b). Study of different h-indices for groups of authors. Journal of the American Society for Information Science and Technology.

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

Egghe, L., & Rousseau, R. (2006b). Systems without low productive sources. Information Processing and Management, 42(6), 1428–1441.

Glänzel, W. (2006a). On the opportunities and limitations of the h-index. Science Focus, 1(1), 10–11, 2006 (in Chinese).

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

Glänzel, W., & Persson, O. (2005). H-index for Price medallists. ISSI Newsletter, 1(4), 15–18.

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.

Hirsch, J.E. (2007). Does the h-index have predictive power ? arXiv:0708.0646v1 [physics.soc-ph], 6 August.

Miller, C.W. (2006). Superiority of the h-index over the impact factor for physics. arXiv:physics/0608183v1, 17 August.

Oppenheim, C. (2007). Using the h-index to rank influential British researchers in information science and librarianship. Journal of the American Society for Information Science and Technology, 58(2), 297–301, 2007.

Prathap, G. (2006). Hirsch-type indices for ranking institutions' scientific research output. Current Science, 91(11), 1439.

Ravichandra Rao, I.K., & Rousseau, R. (2007). Empirical series of journal h-indices. Preprint.

DOI: 10.1002/asi

Rousseau, R. (1992). Concentration and diversity of availability and use in information systems: a positive reinforcement model. Journal of the American Society for Information Science, 43(5), 391–395.

Rousseau, R. (2007a). The influence of missing publications on the Hirsch index. Journal of Informetrics, 1(1), 2–7.

Rousseau, R. (2007b). A case study: Evolution of JASIS' Hirsch index. Preprint.

Saad, G. (2006). Exploring the h-index at the author and journal levels using bibliometric data of productive consumer scholars and business-related journals respectively. Scientometrics 69(1), 117–120.

Schreiber, M. (2007). The influence of self-citation corrections on Egghe's g-index. arXiv:0707.4577v1[physics.soc-ph], 31 July.

Schubert, A. (2007). Successive h-indices. Scientometrics 70(1), 201–205.

Schubert, A. & Glänzel, W. (2007). A systematic analysis of Hirsch-type indices for journals. Journal of Informetrics, 1(3), 179–184.

The STIMULATE6 Group (2007). The Hirsch index applied to topics of interest to developing countries. First Monday, 12(2). Available at: http://www.firstmonday.org/issues/issues/issues12 2/stimulate/

Tol, R.S.J. (2007). A rational, successive g-index applied to economics departments in Ireland. Journal of Informetrics, to appear.

van Raan, A.F.J. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgement for 147 chemistry research groups. Scientometrics, 67(3), 491–502.

Vanclay, J.K. (2007). On the robustness of the h-index. Journal of the American Society for Information Science and Technology, 58(10), 1547–1550.

Vinkler, P. (2007). Eminence of scientists in the light of the h-index and other scientometric indicators. Journal of Information Science, 33(4), 481–491.

Wan, J., Hua, P., & Rousseau, R. (2007). The pure h-index: calculating an author's h-index by taking co-authors into account. Preprint.

Appendix A: Proof of Theorem 2.1

1. Using Equation 13

Clearly, by Equation 13, the analogously defining equation for h^* is

$$\int_{h^*}^{\infty} f^*(j^{*\prime})dj^{*\prime} = h^*$$

Using Equation 4 yields

$$\int_{j^{*\prime}=h^*}^{\infty} f(j')\frac{\psi'(F^{-1}(j'))}{\varphi'(j')}dj^{*\prime} = h^*$$

with $j^{*\prime} = \varphi(j')$, hence

$$\int_{j'=\varphi^{-1}(h*)}^{\infty} f(j')\psi'(F^{-1}(j'))dj' = h^*$$

$$\int_{j'=\varphi^{-1}(h*)}^{\infty} f(j')\psi'(F^{-1}(j'))F'(F^{-1}(j'))d[F^{-1}(j')] = h^*$$

$$\int_{k'=F^{-1}(\varphi^{-1}(h*))}^{\infty} f(F(k'))\psi'(k')F'(k')dk' = h^*$$

where $k' = F^{-1}(j')$, hence $j' = F(k')$. Finally:

$$\int_{k'=F^{-1}(\varphi^{-1}(h*))}^{\infty} f(F(k'))\psi'(k')d[F(k')] = h^*$$

$$\int_{j'=\varphi^{-1}(h*)}^{\infty} f(j')\psi'(F^{-1}(j'))dj' = h^*$$

This is the defining equation for h* in terms of the size-frequency function f, the rank-frequency function F and the transformations φ and ψ.

2. Using Equation 12

For the transformed IPP, we have the analogue of Equation 12: h* is defined as

$$F^*(h^*) = h^*$$

which is, in terms of the functions of the original IPP and of the transformations, by Equation 3:

$$h^* = \varphi(F(\psi^{-1}(h^*))) \tag{A1}$$

(note that, by Equation 1, one is tempted to write $\psi(h) = h^*$ but this is not true if h is the h-index of the original IPP – see further). Alternatively one can also use

$$\varphi^{-1}(h^*) = F(\psi^{-1}(h^*)) \tag{A2}$$

Appendix B: Proof of Theorem 2.2

1. Using Equation 19

Note that

$$\varphi^{-1}(h^*) = j = \left(\frac{h^*}{B}\right)^{\frac{1}{c}}$$

$$F^{-1}(j) = \left(\frac{E}{j}\right)^{\frac{1}{\beta}}$$

and

$$\psi'(r) = Abr^{b-1}$$

This gives in Equation 19, using also Equation 5:

$$\int_{j'=\left(\frac{h*}{B}\right)^{\frac{1}{c}}}^{\infty} \frac{C}{j'^{\alpha}}Ab\left[\left(\frac{E}{j'}\right)^{\frac{1}{\beta}}\right]^{b-1}dj' = h^*$$

$$CAbE^{\frac{b-1}{\beta}}\int_{j'=\left(\frac{h^*}{B}\right)^{\frac{1}{c}}}^{\infty}\frac{dj'}{j'^{\alpha+\frac{b-1}{\beta}}} = h^*$$

Hence (since $\alpha + \frac{b-1}{\beta} > 1$ since $\alpha > 1$ and $b > 1$)

$$\frac{CAbE^{\frac{b-1}{\beta}}}{\alpha + \frac{b-1}{\beta} - 1}\left(\frac{h^*}{B}\right)^{\frac{1-\alpha-\frac{b-1}{\beta}}{c}} = h^*$$

Hence

$$h^{*1-\frac{1-\alpha-\frac{b-1}{\beta}}{c}} = \frac{CAbE^{\frac{b-1}{\beta}}}{B^{\frac{1-\alpha-\frac{b-1}{\beta}}{c}}\left(\alpha + \frac{b-1}{\beta} - 1\right)} \tag{B1}$$

DOI: 10.1002/asi

Note that

$$1-\frac{1-\alpha-\frac{b-1}{\beta}}{c}=\frac{c+b(\alpha-1)}{c}=\delta$$

by Equation 11 and by notation Equation 8. Further

$$\frac{\mathrm{CAbE}^{\frac{b-1}{\beta}}}{\mathrm{B}^{\frac{1-\alpha-\frac{b-1}{\beta}}{c}}\left(\alpha+\frac{b-1}{\beta}-1\right)}=\frac{\mathrm{CAbE}^{(b-1)(\alpha-1)}\beta^{\delta-1}}{c(\delta-1)}$$

Hence, we refound parameter D (see Equation 7): Equation B1 yields

$$h^{*\delta}=\frac{D}{\delta-1}$$

But

$$T^{*}=\int_{B}^{\infty}\frac{D}{j^{*\delta}}dj^{*}=\frac{D}{\delta-1}B^{1-\delta} \tag{B2}$$

(since $j^{*}=\varphi(j)=Bj^{c}\geq B$ since $j\geq 1$). So

$$h^{*\delta}=\frac{T^{*}}{B^{1-\delta}}$$

Hence

$$h^{*}=B^{\frac{\delta-1}{\delta}}T^{*\frac{1}{\delta}}$$

An equivalent calculation would be: follow the calculation of Equations 6, 7, 8 in Egghe (2006a) and apply (23) with T* and using (B2).

We will now show how we can use Equation A1 (or Equation A2).

2. Using Equation A2

Note that

$$\varphi^{-1}(h^{*})=\left(\frac{h^{*}}{B}\right)^{\frac{1}{c}}$$

$$\psi^{-1}(h^{*})=\left(\frac{h^{*}}{A}\right)^{\frac{1}{b}}$$

and that F satisfies Equation 9 with β as in Equation 11. Hence Equation A2 gives

$$\left(\frac{h^{*}}{B}\right)^{\frac{1}{c}}=\frac{E}{\left(\frac{h^{*}}{A}\right)^{\frac{\beta}{b}}} \tag{B3}$$

Use Equation 10 to get

$$E=\left(\frac{C}{\alpha-1}\right)^{\frac{1}{\alpha-1}}=T^{\frac{1}{\alpha-1}} \tag{B4}$$

since

$$T=\int_{1}^{\infty}f(j)dj=\int_{1}^{\infty}\frac{C}{j^{\alpha}}dj=\frac{C}{\alpha-1}$$

since $\alpha>1$, T being the total number of sources. But

$$T^{*}=\psi(T)=AT^{b} \tag{B5}$$

since ψ increases. Equations B5 and B4 in Equation B3 yields

$$h^{*}=B^{\frac{\frac{1}{c}}{\frac{1}{c}+\frac{\beta}{b}}}T^{*\frac{\frac{1}{b(\alpha-1)}}{\frac{1}{c}+\frac{\beta}{b}}}$$

which is readily seen to be the same as Equation 22, using Equation 11.

Appendix C: Proof of Theorem 2.3

We have that

$$\psi'(r)=Abr^{b-1}$$

$$\psi^{-1}(g^{*})=\left(\frac{g^{*}}{A}\right)^{\frac{1}{b}}=r$$

Hence we have the equation, based on Equation 25

$$\int_{0}^{\left(\frac{g^{*}}{A}\right)^{\frac{1}{b}}}B(F(r))^{c}Abr^{b-1}dr=g^{*^{2}}$$

Using Equation 9 gives

$$E^{c}BAb\int_{0}^{\left(\frac{g^{*}}{A}\right)^{\frac{1}{b}}}r^{b-1-\beta c}dr=g^{*2}$$

For $\beta c+1-b<1$ we have

$$\frac{E^{c}BbA^{\frac{\beta c}{b}}}{b-\beta c}g^{*\frac{b-\beta c}{b}}=g^{*2}$$

Hence

$$g^{*1+\frac{\beta c}{b}}=\frac{BbA^{\frac{\beta c}{b}}E^{c}}{b-\beta c}$$

Now, we use Equations 11 and B5 yielding

$$g^{*}=\left(\frac{Bb(\alpha-1)}{b(\alpha-1)-c}\right)^{\frac{b(\alpha-1)}{b(\alpha-1)+c}}T^{*\frac{c}{b(\alpha-1)+c}}$$

hence, using notation Equation 8 for δ we find

$$g^{*}=B^{\frac{\delta-1}{\delta}}T^{*\frac{1}{\delta}}\left(\frac{\delta-1}{\delta-2}\right)^{\frac{\delta-1}{\delta}}$$

DOI: 10.1002/asi

Appendix D: Proof of Theorem 3.2

(i) Let $h^* = \psi(h)$. Then, by definition of the h-index in both systems and by Lemma 3.1 we have

$$\begin{aligned} h^* &= F^*(h^*) = F^*(\psi(h)) \\ &= \varphi(F(h)) = \varphi(h) \end{aligned}$$

proving Equation 28.

(ii) Let $h^* < \psi(h)$. Then, since F^* is defined on h^* and $\psi(h)$ (both belonging to $[0, T^*]$) and since F^* strictly decreases, we have

$$h^* = F^*(h^*) > F^*(\psi(h))$$

Again invoking the above lemma yields

$$h^* > \varphi(h)$$

hence proving Equation 29.

(iii) Let $h^* > \psi(h)$. By the same argument we have

$$h^* = F^*(h^*) < F^*(\psi(h))$$

Hence by the above Lemma:

$$h^* < \varphi(h)$$

proving Equation 30. □

DOI: 10.1002/asi