Performance and its relation with productivity in Lotkaian systems

by

L. Egghe

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

ABSTRACT

In general information production processes (IPPs), we define productivity as the total number of sources but we present a choice of seven possible definitions of performance: the mean or median number of items per source, the fraction of sources with a certain minimum number of items, the h-, g-, R- and h_w-index. We give an overview of the literature on different types of IPPs and each time we interpret "performance" in these concrete cases. Examples are found in informetrics (including webometrics and scientometrics), linguistics, econometrics and demography.

In Lotkaian IPPs we study these interpretations of "performance" in function of the productivity in these IPPs. We show that the mean and median number of items per source as well as the fraction of sources with a certain minimum number of items are increasing

---

[1] Permanent address
Acknowledgement: The author thanks prof. Dr. P. Bonckaert for discussions leading to condition (iii) in Proposition IV.2.
Key words and phrases: Lotka, productivity, performance, information production process, IPP, h-index, g-index, R-index, h_w-index, mean, median

functions of the productivity if and only if the Lotkaian exponent is decreasing in function of the productivity. We show that this property implies that the g-, R- and h_w-indices are increasing functions of the productivity and, finally, we show that this property implies that the h-index is an increasing function of productivity. We conclude that the h-index is the indicator which shows best the increasing relation between productivity and performance.

I. Introduction

Suppose we fix one author, as a first example. This author has published a certain number of papers which we will denote by T. The number T, clearly, can be considered as the productivity of this author. Next we look at the citations to these papers. In this situation we can wonder how to describe the "performance" of this author with respect to this publication-citation situation. In this example, performance could be interpreted as visibility of the author's papers.

In this introduction we will present several ways to express performance. Perhaps the simplest way is defining performance (in the above example) as the average number of citations per paper: $\mu = \frac{A}{T}$ where A denotes the total number of citations to the T papers of this author. Another way of expressing performance is by replacing the average μ by the median Md of the citation data. Both methods are similar but, of course, not equivalent. They both use the citation data itself but in a different way: μ uses the sum of all citations while Md is this number of citations (possibly interpollated) such that 50% of the papers have more than (or equal to) this number of citations. In this context we could even replace Md by any percentile P_c where c expresses the used percentage $c \in \,]0,100[$. So, in case of Md (or P_c), we do not even use all citation data (or at least, citation data are used indirectly).

A very different way of defining performance is by looking at the fraction of papers with at least a certain number (denoted a) of citations. Here performance is expressed by a fraction of papers (not citations !) and where again the citation data are used indirectly.

The last two methods (the percentile way and the paper fraction way) are ingeniously combined in the h-index (Hirsch (2005)) defined as follows. Arrange the papers of this author decreasingly according to their number of received citations. The h-index is the largest rank h such that all papers on ranks $1, 2, ..., h$ have at least h citations. This remarkable definition has become a real hype. Defined only in 2005, the h-index has been studied in numerous papers (in and outside the field of informetrics), is available in citation databases such as Scopus, Web of Science (WoS) and on the Google Scholar based website “Publish or Perish” at www. harzing.com/pop.htm and will become a hiring and evaluation tool for scientist. It is clear that the h-index is a measure of performance (e.g. of an author’s career) where, again, the citation data are only indirectly (and certainly only partially) used. The latter fact is a negative point for the h-index (its most positive point being its simplicity) since, once an article belongs to the first h ones (the so-called h-core, but I do not like the term core here since these h articles are not really a core: they are used to define the number h) it does not matter anymore how much more (than h) citations it has (or will) receive(d).

To overcome this problem, Egghe defined in Egghe (2006) the so-called g-index. In the same ranking of papers as above, the g-index is the largest rank g such that all papers on ranks $1, 2, ..., g$ have, together, at least g^2 citations. This definition was inspired by the fact that the h-index satisfies this criterion (hence, since g is the largest rank with this property, we have that $g \geq h$ but that is not important in itself) and by the fact that, now, we effectively use the actual number of citations to the papers on ranks $1, ..., g$. Formulated otherwise (see also Schreiber (2007)) the g-index is the largest rank g such that the papers on rank $1, 2, ..., g$ have, <u>on average</u>, g citations. As the h-index, the g-index uses paper ranks as well as citation numbers and, in addition, the g-index uses a citation average. In a way we can say that the g-index is a kind of combination of the first three methods of defining performance: average of citations, percentiles of citations and fraction of papers.

Other indices which also try to improve the h-index on the actual use of the citation data of the most cited papers are the R-index (Jin, Liang, Rousseau and Egghe (2007)) and the weighted h-index h_w (Egghe and Rousseau (2008)). The R-index is the square root of the sum of the citations to the first h papers (hence uses the h-index itself) and the weighted h-index h_w uses the citations to the papers as weights for the paper ranks.

It is clear that all these performance measures can also be defined on other source-item relations (other than paper-citation relations), i.e. to general IPPs. Here we have general sources producing (or having) general items and, as before, we can define the performance measures as we did for the paper-citation IPP. Here the average number μ of items per source can be calculated (first method) or, based on the number of items per source, we can calculate Md (or the percentiles P_c) (second method) or the fraction of sources with at least a items (third method). For the h-, g-, R- and h_w-index we use the ranking of the sources being decreasing in the number of items per source.

In the next section we will study several concrete examples of IPPs, other than the paper-citation one. This will lead to newly defined performance measures not known before in these fields. We will describe papers-authors IPPs leading to new collaboration measures (such as the collaboration h-index), papers-downloads IPPs leading to "use" measures, IPPs dealing with authors as sources (or journals as sources), IPPs consisting of books and their borrowings, IPPs of linguistical texts; econometric IPPs where performance is now wealth, demographic IPPs where performance describes population densities and, finally, social network IPPs such as internet, WWW, intranets, including the already mentioned citation and collaboration networks.

In the third section we will calculate all these performance measures in a Lotkaian framework (or we will refer to the already existing result), i.e. where we have the decreasing power law

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

$C > 0$, $\alpha > 1$ where $f(j)$ is the size-frequency function describing the density of sources $f(j)$ with item density $j \geq 1$, cf. Egghe (2005).

In the fourth section we are interested in the relation between productivity (the total number T of sources) and performance (in its several versions as defined above) in Lotkaian IPPs. Particularly we are interested in the possible increasing relation between productivity and performance. This interest originates from the problem raised in collaboration situations: does collaboration leads to a higher productivity (number of papers), cf. earlier studies Beaver and

Rosen (1979), Bordons and Gómez (2000), Borgman and Furner (2002), Pao (1981, 1982, 1992), Price and Beaver (1966), Subramanyam (1983), Zuckerman (1967) and the recent Egghe, Goovaerts and Kretschmer (2008).

Of course, as described above, we are in a position to study the general performance-productivity relation in general IPPs. We prove necessary and sufficient conditions for having performance as an increasing function of productivity, where all the above defined performance measures are used. Relations between these conditions are presented and we show that the h-type indices versus T (productivity) is increasing in more cases than in case we use μ, Md (or P_c) or the fraction of sources with at least a items. Amongst the h-type indices we show that the h-index itself is an increasing function of T in more cases than in case we use the g-, R- or h_w-index.

The paper closes with a conclusions and open problems section.

II. Performance in diverse IPPs

In this section we will discuss several diverse IPPs, in and outside informetrics, including scientometric, webometric, econometric examples as well as examples in social networks (including internet, WWW, intranets, citation and collaboration networks), linguistics, libraries and even demography.

II.1 papers-citation system

For the sake of completeness we repeat the system which we discussed already in the Introduction. For clarity reasons and also for easy comparisons we will describe each system in a schematic way.

IPP: an author (can also be several authors (e.g. of an institute) but then we consider them as one “meta” author)
In this IPP we have
sources: the papers of this author
items: the citations to these papers

Performance: visibility (based on received citations)

Performance measures: as discussed in the Introduction

(I) μ = average number of citations per paper

(II) Md = median of the number of citations of the papers (extension: percentiles P_c, $c \in]0,100[$)

(III) $\frac{\varphi(a)}{T}$ = fraction of the papers with a or more citations, $a > 1$

(IV) h-type indices (h, g, R, h_w) are the "classical" ones in the publication-citation system. In case we have a meta author, $h = h_G$, the global h-index as discussed in van Raan (2006) and Egghe and Rao (2008).

II.2 papers - co-authors system

IPP: an author

sources: the papers of this author

items: the co-authors of these papers (including the author himself or herself)

Performance: collaboration

Performance measures:

(I) μ = average number of authors per paper

(II) Md = median of the number of authors of the papers (extension: P_c)

(III) $\frac{\varphi(a)}{T}$ = fraction of the papers with a or more authors $(a > 1)$

(IV) The h-type indices (h, g, R, h_w) are collaboration indices (to the best of our knowledge, introduced here for the first time).

This system hence comprises collaboration and, based on the general theory to be developed, will yield results on the well-known problem (see the Introduction and the references given there) of the relation between productivity T and the performance collaboration.

II.3 Papers-downloads system

IPP: an author

sources: the papers of this author

items: the downloads of these papers.

It is clear that we can deal here with electronic articles in an e-journal or in an institutional repository.

Performance: use (or visibility based on downloads)
Performance measures:

(I) μ = average number of downloads per paper

(II) similar for Md or P_c

(III) $\frac{\varphi(a)}{T}$ = fraction of the papers with a or more downloads $(a > 1)$

(IV) The h-type indices are use indices.

II.4 Authors-publications system

IPP: a group of authors (e.g. in an institute or a field)
sources: the authors
items: their publications (cf. the classical paper Lotka (1926))
Performance: paper production of the group (note that this has nothing to see with productivity in the title of this article: as always, the general term productivity in the title of this paper refers to the total number T of sources being here the total number of authors. The term productivity is less appropriate here but due to the specific application; this will also be noted in further applications.
Performance measures:

(I) μ = mean number of papers per author

(II) Md (or P_c) based on the number of papers of each author

(III) $\frac{\varphi(a)}{T}$ = fraction of the authors with a or more papers $(a > 1)$

(IV) The h-type indices are publication indices. For the h-index we introduced it in Egghe and Rao (2008) as h_P.

II.5 Authors-citations system

IPP: a group of authors

sources: the authors

items: their total number (per author) of received citations

Performance: visibility of the group

Performance measures: this is similar to the previous system where we replace "publications" by "citations". Note that in this case, the h-index is $h = h_C$ as introduced in Egghe and Rao (2008).

II.6 Authors – h-index system

IPP: a group of authors

sources: the authors

items: their h-index

Performance: h-index scoring of the group

Performance measures:

(I) μ = average h-index per author

(II) Md (or P_c): similar

(III) $\frac{\varphi(a)}{T}$ = fraction of authors with an h-index $h \geq a$ $(a > 1)$

(IV) The h-type indices are based on these h-indices. For the h-index itself this means $h = h_2$, the successive h-index on second level, introduced (independently) by Prathap (2006) and Schubert (2007) – see also Egghe (2008). Of course, a variant of this system is using another index, e.g. the g-index: then the g-index of this system is the successive g-index g_2.

The above three models, where the IPP consists of a group of authors can be modified by considering a group of journals (e.g. in a field or from a publisher). h-indices of journals where introduced in Braun, Glänzel and Schubert (2005, 2006).

II.7 Books-loans system

IPP: a library

sources: the books in this library

items: their loans (in a certain period)

Performance: loan scores of this library

Performance measures:

(I) μ = average number of loans per book

(II) Md (or P_c): similar

(III) $\frac{\varphi(a)}{T}$ = fraction of the books that where checked out a times of more $(a > 1)$. As always, we only consider the books that were checked out at least once (which is, in this case, probably a small part of the library).

(IV) The h-type indices can characterise certain libraries (e.g. public versus scientific libraries).

II.8 Type-token system (linguistics)

IPP: a text

sources: word types

items: word tokens, i.e. their use in the text (cf. Herdan (1960)).

Performance: the pattern of word use in the text (e.g. as being characteristic for a writer)

Performance measures: similar as in the other cases, introducing here h-type indices for texts.

II.9 Employees-salary system (econometrics)

IPP: a group of employees (e.g. in a company)

sources: the employees

items: their salaries (variant: their produced items)

Performance: wealth

Performance measures: similar as in the other cases, introducing here h-type indices in econometrics.

For an econometric reference, dealing with similar distributions as we do in informetrics, we refer the reader to Lambert (2001).

II.10 Cities-inhabitants system

IPP: a community (such as, e.g., a country)
sources: cities and villages
items: their inhabitants
Performance: degree of habitation
Performance measures: similar as above, now introducing h-type indices for communities

For more on this, see Marsili and Zhang (1998), Ioannides and Overman (2003), Batty (2003) but even Zipf (1949) mentions this application. Variants: sizes of firms: Axtell (2001), sizes of domain names in the internet: Rousseau (1997).

II.11 Nodes-links system

IPP: a (social) network (includes the first two systems as well: citation and collaboration network but here we will give a web example)
sources: nodes (e.g. web sites)
items : links (in- or out-links)
performance: visibility
Performance measures: as above now introducing h-type indices for networks, based on hyperlinks.

Many references on this application: Bilke and Peterson (2001), Jeong, Tombor, Albert, Ottval and Barabási (2000), Barabási, Jeong, Néda, Ravasz, Schubert and Vicsek (2002), Adamic, Lukose, Puniyani and Huberman (2001), Barabási and Albert (1999), Adamic and Huberman (2001, 2002), Bornholdt and Ebel (2001), Albert, Jeong and Barabási (1999), Barabási, Albert and Jeong (2000), Thelwall and Wilkinson (2003), Faloutsos, Faloutsos and Faloutsos (1999), Rousseau (1997) – see also the excellent books Pastor-Satorras and Vespignani (2004) and Huberman (2001).

II.12 Nodes-accesses system

IPP: a (social) network (it also contains the third example)
sources: nodes (e.g. web sites)
items: accesses (or page views)

Performance: use in this sense

Performance measures: as above now introducing a general use-h (-type) index.

Further reading: Nielsen (1997), Huberman, Pirolli, Pitkow and Lukose (1998), Aida, Takahashi and Abe (1998) – see also Huberman (2001) on "the law of surfing". Variant: replace "accesses" by "visitors" – see Adamic and Huberman (2001, 2002).

II.13 Websites content

IPP: a group of websites

sources: websites

items: their web pages

Performance: content

Performance measures: as above, now introducing content -h (-type) indices.

See also Adamic and Huberman (2001, 2002).

III. Performance measures in a Lotkaian IPP

Let us suppose we have an IPP with size-frequency function

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

$C > 0,\ \alpha > 1,\ j \geq 1$. In this section we will present concrete formulae for the four types of performance measures which were introduced in the previous sections.

III.1 The average μ

The formula can be found in Egghe (2005) but we recalculate it here since the proof is only two lines. We have, for $\alpha > 2$

$$\mu = \frac{A}{T} = \frac{\alpha - 1}{\alpha - 2} \tag{3}$$

where A denotes the total number of items, being

$$A = \int_1^{\infty} j f(j) dj = \frac{C}{\alpha - 2} \tag{4}$$

and where T denotes the total number of sources, being

$$T = \int_1^{\infty} f(j) dj = \frac{C}{\alpha - 1} \tag{5}$$

from which (3) readily follows. Note that the requirement $\alpha > 2$ is needed for the convergence of the integral in (4). If we limit j to an upper limit (which we do not here for the sake of simplicity) we can suffice by taking $\alpha > 1$.

III.2 The median Md and percentiles P_c

The percentile P_c, for $c \in]0,100[$ is this item density for which we have that c% of the sources have item density j or higher (classically in statistics, P_c is defined using 100-c instead of c above but, since we want to emphasize on the higher item densities we define P_c as given above). For c=25, 50 and 75 we have the three quartiles and for c=50, P_c is also called the median, denoted by Md. By definition of P_c we have, for $\alpha > 1$

$$\int_{P_c}^{\infty} f(j) dj = \frac{c}{100} \int_1^{\infty} f(j) dj \tag{6}$$

By (1), (6) yields

$$\frac{C}{1-\alpha} P_c^{1-\alpha} = \frac{c}{100} \frac{C}{\alpha - 1}$$

$$P_c = \left(\frac{c}{100} \right)^{\frac{1}{1-\alpha}}$$

$$P_c = \left(\frac{100}{c}\right)^{\frac{1}{\alpha-1}} \qquad (7)$$

which for c=50 yields

$$P_{50} = Md = 2^{\frac{1}{\alpha-1}} \qquad (8)$$

Note that the calculation of all P_c values is possible for all $\alpha > 1$. Although this calculation is simple, we cannot find a reference for it.

III.3 The fraction $\frac{\varphi(a)}{T}$ of the sources with a or more items (densities), a>1

By definition this fraction is

$$\frac{\varphi(a)}{T} = \frac{\int_a^\infty f(j)dj}{\int_1^\infty f(j)dj} \qquad (9)$$

Here $\varphi(a)$ is the number of sources with a or more items (densities). This looks similar to (6) but note that P_c is an item density while $\frac{\varphi(a)}{T}$ is a fraction of sources, hence a completely different measure. We have, by definition

$$\frac{\varphi(a)}{T} = \frac{\frac{C}{\alpha-1}\frac{1}{a^{\alpha-1}}}{\frac{C}{\alpha-1}}$$

$$\frac{\varphi(a)}{T} = \frac{1}{a^{\alpha-1}} \qquad (10)$$

Finally, we present formulae for the h-type indices h, g, R and h_w. These formulae are less trivial but where published earlier. We will give these formulae without proof but present the right reference.

III.4 The h-type indices h, g, R and h_w

In Egghe and Rousseau (2006) we proved, for $\alpha > 1$

$$h = T^{\frac{1}{\alpha}} \tag{11}$$

In Egghe (2006) we proved, for $\alpha > 2$ that

$$g = \left(\frac{\alpha - 1}{\alpha - 2}\right)^{\frac{\alpha - 1}{\alpha}} T^{\frac{1}{\alpha}} \tag{12}$$

or

$$g = \left(\frac{\alpha - 1}{\alpha - 2}\right)^{\frac{\alpha - 1}{\alpha}} h \tag{13}$$

In Jin, Liang, Rousseau and Egghe (2007) we proved, for $\alpha > 2$ that

$$R = \left(\frac{\alpha - 1}{\alpha - 2}\right)^{\frac{1}{2}} T^{\frac{1}{\alpha}} \tag{14}$$

$$R = \left(\frac{\alpha - 1}{\alpha - 2}\right)^{\frac{1}{2}} h \tag{15}$$

Finally, In Egghe and Rousseau (2008) we proved that, for $\alpha > 2$

$$h_w = \left(\frac{\alpha - 1}{\alpha - 2}\right)^{\frac{1}{2(\alpha - 1)}} T^{\frac{1}{\alpha}} \tag{16}$$

$$h_w = \left(\frac{\alpha - 1}{\alpha - 2}\right)^{\frac{1}{\alpha(\alpha-1)}} h \tag{17}$$

This concludes the presentation of the performance measures. In the next section we are interested in the relation between these performance measures and the productivity, defined as T, the total number of sources.

Remark:
In earlier publications (e.g. Egghe (2005)) we also used the terminology that "sources produce or have items". So this is also a kind of productivity which terminology, in order not to confuse with the total number of sources, is not used here. It is exactly the source-item relation that forms the basis for our study of performance. The sources are produced by the IPP (e.g. an author (=IPP) produces articles which performance is studied, e.g. based on citations received by these articles – but see also the many other examples in the second section).

IV. The relation between performance and productivity

In general Lotkaian systems we do not have that performance (in any definition) is increasing with productivity T. We start this section by giving the necessary counterexamples.

First example: $A = 100{,}000$, $T = 2{,}000$, hence $\mu = 50$. By Proposition II.2.1.1.1 in Egghe (2005) we have an existing Lotkaian system where C and α (formula (1)) are given by

$$\alpha = \frac{2A - T}{A - T} \tag{18}$$

$$C = \frac{AT}{A - T} \tag{19}$$

Hence $\alpha = 2.0204$ and $C = 2{,}041$.

Second example: $A = 100{,}000$, $T = 4{,}000$, hence $\mu = 25$. Using Proposition II.2.1.1.1 in Egghe (2005) again we have an existing Lotkaian system where C and α are (use (18) and (19)) $\alpha = 2.0417$ and $C = 4{,}166$.

Comparing both Lotkaian examples we see that, in the second example, T is larger than in the first one while μ is smaller in the second example, when compared with the first one. Hence, in general μ is not an increasing function of T.

Here μ was calculated as $\frac{A}{T}$ but the result also follows if we use formula (3), as is readily seen.

Since α has increased from $\alpha = 2.0204$ in the first example to $\alpha = 2.0417$ in the second one, we also see from (7) and (8) that all P_c values (including Md) have decreased too (while productivity T has increased). The same examples can be used for the fraction $\frac{\varphi(a)}{T}$ of sources with a or more items (densities). This is readily seen by (10) since a is a constant.

For counterexamples on the general increasing relation between h and T, g and T, R and T and h_w and T, we will use the first example and a third one which is now given.

Third example: $A = 10{,}000$, $T = 2{,}500$, hence we have an existing Lotkaian system if we take, by (18) and (19), $\alpha = 2.3333...$ and $C = 3333.33...$

For the first example we have, by (10): $h = 43.037767$ while for the third example we have $h = 28.59302$, hence decreased (while T has increased: $2{,}000$ in the first example and $2{,}500$ in the third one). Similarly, both examples can also be used to produce counterexamples in the case of g, R and h_w: first example: $g = 310.45452$, $R = 304.38263$, $h_w = 292.70834$; third example: $g = 63.13851$, $R = 57.186042$, $h_w = 48.087536$, yielding the requested counterexamples.

We will now prove some necessary and sufficient conditions on when the performance measures are increasing functions of the productivity T.

Since we study the relation with T we will denote the performance measures as function of T: $\mu(T)$, and so on.

Proposition IV.1: The following assertions are equivalent:

(i) $\mu(T)$ increases (strictly) in T

(ii) $P_c(T)$ increases (strictly) in T for all $c \in]0,100[$ (hence including P_{50}=Md)

(iii) $\frac{\varphi(a)(T)}{T}$ increases (strictly) in T

(iv) $\alpha(T)$ decreases (strictly) in T

Proof: It is readily seen that (3) is a decreasing function of α. The same is true for (7) for all $c \in]0,100[$ and for (10) for all $a > 1$. Hence the result follows (with or without the term "strictly"). □

Note that condition (iv) is equivalent with the (strict) increase of the Lorenz-curve of the rank-frequency function of the IPP: see Egghe (2005), Corollary IV.3.2.1.5. Since we do not use this result further on we will not go into this in more detail: these can be found in Egghe (2005), p. 204-205.

For $h(T)$, we have by (10):

$$h(T) = T^{\frac{1}{\alpha(T)}} \tag{20}$$

We immediately see that, under the conditions of Proposition IV.1 we have, by (20), that $h(T)$ is an increasing function of T. But the condition is not necessary: the next proposition shows that $h(T)$ increases in T in a lot more cases.

Proposition IV.2: The following assertions are equivalent, if $T > 1$:

(i) $h(t)$ increases strictly in T

(ii)

$$\alpha'(T) < \frac{\alpha(T)}{T \ln T} \tag{21}$$

(iii) $\frac{\alpha(T)}{\ln T}$ strictly decreases in T. This is, of course, satisfied if $\alpha'(T) < 0$ (supposing, evidently, $T > 1$).

Proof: The function $h(T) = T^{\frac{1}{\alpha(T)}}$ is a power function in combination with an exponential function. If we denote this symbolically by u^v we have, as is well-known

$$(u^v)' = vu^{v-1}u' + u^v(\ln u)v' \tag{22}$$

which will be applied to $h(t)$:

$$h'(T) = \frac{1}{\alpha(T)} T^{\frac{1}{\alpha(T)}-1} - \frac{\alpha'(T)}{\alpha^2(T)}(\ln T) T^{\frac{1}{\alpha(T)}}$$

$$> 0$$

if and only if (21) is valid, since $T > 1$. The equivalence of (i) and (ii) is hence proved. That (ii) and (iii) are equivalent follows: $\frac{\alpha(T)}{\ln T}$ decreases strictly in T if and only if its derivative is strictly negative. This gives the equivalent condition

$$\frac{(\ln T)\alpha'(T) - \frac{\alpha(T)}{T}}{(\ln T)^2} < 0$$

whence (21) since the denominator is positive and since $T > 1$. □

A strict upper limit for $\alpha(T)$ is given by $\ln T$ (or $b \ln T$ for all $b > 0$ constant) since for this function, (21) becomes an equality.

The next example shows that condition (21) is much weaker than the condition $\alpha'(T) < 0$.

<u>Example</u>: We present a strictly increasing function $\alpha(T)$ (hence for which, by Proposition IV.1, $\mu(T)$, $P_c(T)$ and $\frac{\varphi(a)(T)}{T}$ are not increasing; in fact, by the formulae for these measures, these measures are strictly decreasing) for which the function $h(T)$ strictly increases. Based on the above remark, we take, for $T > 1$

$$\alpha(T) = \ln^{\varepsilon}(T) \tag{23}$$

with $\varepsilon \in]0,1[$ fixed. Then

$$\alpha'(T) = \frac{\varepsilon}{T \ln^{1-\varepsilon}(T)} > 0$$

But

$$\frac{\alpha(T)}{T \ln T} = \frac{\ln^{\varepsilon}(T)}{T \ln T}$$

$$= \frac{1}{T \ln^{1-\varepsilon}(T)},$$

hence (21) is satisfied since $\varepsilon \in]0,1[$ and since we take $T > 1$.

Alternatively, one could remark that $\frac{\alpha(T)}{\ln T}$ strictly decreases so that condition (iii) in Proposition IV.2 is satisfied, hence also (21).

Important note: In cases of growing IPPs, e.g. in the case of the evolving career of an author, we know, trivially, that $h(T)$ is increasing in T. As a corollary of this, the evolution of this IPP, involving a T-dependent $\alpha(T)$, always satisfies (21), which is a new result in Lotkaian informetrics:

Corollary IV.3: For growing Lotkaian IPPs we have that the evolution of the Lotka exponent $\alpha(T)$ is limited to the condition

$$\alpha'(T) < \frac{\alpha(T)}{T \ln T}$$

Note that this is a result obtained from h-index theory without referring to this theory anymore.

From Proposition IV.1 and IV.2 we have the corollary, if $T > 1$:

Corollary IV.4: The equivalent conditions (i)-(iv) in Proposition IV.1, e.g. $\mu(T)$ increases in T, imply $h(T)$ increases in T (for $T > 1$) but not vice-versa.

This means, in practise, that $h(T)$ is, in more cases than $\mu(T)$, $P_c(T)$ or $\frac{\varphi(a)(T)}{T}$, increasing in T. This needs further practical investigation.

We now turn our attention to the other h-type indices.

Proposition IV.5: The following assertions are equivalent:

(i) $g(T)$ increases strictly in T

(ii)

$$\alpha'(T)\left(\frac{\ln T}{\alpha(T)} + \frac{1}{\alpha(T) - 2} - \frac{1}{\alpha(T)} \ln\left(\frac{\alpha(T) - 1}{\alpha(T) - 2}\right)\right) < \frac{1}{T} \tag{24}$$

Proof: The product rule for derivatives and (22) imply on (12), after some calculation, that $g'(T) > 0$ if and only if (24) is valid. We leave the relatively long calculation to the reader or the proof can always be obtained from the author. □

The above result has an interesting consequence.

Corollary IV.6: We have that (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) but not vice-versa: for $T > 1$:

(i) The equivalent conditions of Proposition IV.1: e.g. $\mu(T)$ increases strictly in T

(ii) $g(T)$ increases strictly in T

(iii) $h(T)$ increases strictly in T.

Proof: the proof is finished if we can show that

$$\frac{1}{\alpha(T)-2} - \frac{1}{\alpha(T)} \ln\left(\frac{\alpha(T)-1}{\alpha(T)-2}\right) > 0 \tag{25}$$

But

$$\frac{\alpha(T)}{\alpha(T)-2} > \frac{\alpha(T)-1}{\alpha(T)-2} > \ln\left(\frac{\alpha(T)-1}{\alpha(T)-2}\right)$$

trivially, hence (25) is proved. From this it follows that the expression between brackets in (24) is positive, hence the condition $\alpha'(T) < 0$ implies (24). This proves (i) $\Rightarrow$ (ii) by Proposition IV.1. Furthermore since (25) is valid, (24) implies

$$\alpha'(t)\frac{\ln T}{\alpha(T)} < \frac{1}{T}$$

which is condition (21), hence (ii) $\Rightarrow$ (iii) is proved. □

Similar results are obtained for the indices R and h_w.

Proposition IV.7: The following assertions are equivalent:

(i) $R(T)$ increases strictly in T

(ii)

$$\alpha'(T)\left(\frac{\ln T}{\alpha(T)}+\frac{\alpha(T)}{2(\alpha(T)-2)(\alpha(T)-1)}\right)<\frac{1}{T} \tag{26}$$

Proof: Again the proof follows from the product rule for derivatives, (22) and (14) yielding $R'(T)>0$ if and only if (26) is valid, hence the equivalence of (i) and (ii). Again, the calculations are left to the reader or can be obtained from the author. □

A similar corollary as Corollary IV.6 can be proved here.

Corollary IV.8: We have that (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) but not vice-versa: for $T>1$:

(i) The equivalent conditions of Proposition IV.1: e.g. $\mu(T)$ increases strictly in T

(ii) $R(T)$ increases strictly in T

(iii) $h(T)$ increases strictly in T.

Proof : Clearly, in (26), the expression

$$\frac{\alpha(T)}{2(\alpha(T)-2)(\alpha(T)-1)}>0 \tag{27}$$

(since $\alpha(T)>2$). So, the expression between brackets in (26) is positive so that the condition $\alpha'(T)<0$ implies (26). This proves (i) $\Rightarrow$ (ii) by Proposition IV.1. Furthermore, since (27) is valid, (26) implies

$$\alpha(T)\frac{\ln T}{\alpha(T)}<\frac{1}{T}$$

which is condition (21), hence (ii) $\Rightarrow$ (iii) is proved. □

Finally for h_w we have the following result.

Proposition IV.9: The following assertions are equivalent:

(i) $h_w(T)$ increases strictly in T

(ii)

$$\alpha'(T)\left(\frac{\ln T}{\alpha(T)}+\frac{\alpha(T)}{2(\alpha(T)-1)^2}\ln\left(\frac{\alpha(T)-1}{\alpha(T)-2}\right)+\frac{\alpha(T)}{2(\alpha(T)-1)^2(\alpha(T)-2)}\right)<\frac{1}{T} \tag{28}$$

Proof: Again the proof follows from the product rule for derivatives, (22) and (16) yielding $h'_w(T)>0$ if and only if (28) is valid, hence the equivalence of (i) and (ii). Again the calculations are left to the reader or can be obtained from the author. □

We also have the Corollary IV.10, as we had for $g(T)$ and $R(T)$.

Corollary IV.10: We have that (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) and not vice-versa: for $T>1$:

(i) The equivalent conditions of Proposition IV.1: e.g. $\mu(T)$ increases strictly in T

(ii) $h_w(T)$ increases strictly in T

(iii) $h(T)$ increases strictly in T.

Proof: The proof follows the lines of Corollaries IV.6 and IV.8, now remarking that

$$\frac{\alpha(T)}{2(\alpha(T)-1)^2}\ln\left(\frac{\alpha(T)-1}{\alpha(T)-2}\right)+\frac{\alpha(T)}{2(\alpha(T)-1)^2(\alpha(T)-2)}>0$$

since $\alpha(T)>2$. □

A relation between the assertions of Propositions IV.5, IV.7 and IV.9 is not clear.

V. Conclusions and open problems

Four different types of "performance" of an IPP have been given and interpreted in several IPP examples, hereby introducing also new h-type indices.

Formulae for these performance measures have been given in the context of Lotkaian informetrics.

The important possible relation between "performance" and productivity (expressed as the total number T of sources in the IPP) has been investigated. Necessary and sufficient conditions for an increasing functional relation between performance (in all its defined versions) and productivity have been given. We showed, from these results that the mean, median (more generally: all percentiles), average fraction of sources with at least a certain number of items are strictly increasing functions of T iff $\alpha(T)$ is strictly decreasing in T. These conditions imply $g(T)$, $R(T)$ and $h_w(T)$ to be strictly increasing in T and each of these conditions imply that $h(T)$ is an increasing function of T.

Thus $h(T)$ increases in T in more cases than all the other performance measures which is an interesting feature of the h-index with respect to the establishment of the relation between performance and productivity.

We intend to examine these results in some practical cases and invite the reader to do so as well. Especially since there are so many, very different examples of IPPs as indicated in this paper, many different intriguing experiments can be set-up (in completely different environments) to discover the relation between performance and productivity. Evidently, it is also a challenge to define other performance measures in an IPP and to relate them with productivity.

References

L.A. ADAMIC and B.A. HUBERMAN (2001). The Web's hidden order. Communications of the ACM 44(9), 55-60.

L.A. ADAMIC and B. HUBERMAN (2002). Zipf's law and the Internet. Glottometrics 3, 143-150.

L.A. ADAMIC, R.M. LUKOSE, A.R. PUNIYANI and B.A. HUBERMAN (2001). Search in power-law networks. Physical Review E 64, 46135-46143.

M. AIDA, N. TAKAHASHI and T. ABE (1998). A proposal of dual Zipfian model for describing HTTP access trends and its application to address cache design. IEICE Transactions on Communication E81-B(7), 1475-1485.

R. ALBERT, H. JEONG and A.-L. BARABÁSI (1999). Diameter of the World-Wide Web. Nature 401, 130-131.

R.L. AXTELL (2001). Zipf distribution of U.S. firm sizes. Science 293, 1818-1820.

A.-L. BARABÁSI and R. ALBERT (1999). Emergence of scaling in random networks. Science 286, 509-512.

A.-L. BARABÁSI, R. ALBERT and H. JEONG (2000). Scale-free characteristics of random networks: the topology of the world-wide-web. Physica A 281, 69-77.

A.-L. BARABÁSI, H. JEONG, Z. NÉDA, E. RAVASZ, A. SCHUBERT and T. VICSEK (2002). Evolution of the social network of scientific collaborations. Physica A 311, 590-614.

M. BATTY (2003). The emergence of cities: complexity and urban dynamics. http://www.casa.ucl.ac.uk/working_papers/paper64.pdf

D. DEB. BEAVER and R. ROSEN (1979). Studies in scientific collaboration. Part II. Scientific co-authorship, research productivity and visibility in the French scientific elite, 1799-1830. Scientometrics 1(2), 133-149.

S. BILKE and C. PETERSON (2001). Topological properties of citation and metabolic networks. Physical Review E 6403(3), 76-80.

M. BORDONS and I. GÓMEZ (2000).Collaboration networks in science. Festschrift in honor of E. Garfield, ASIST, Chapter 10, 197-213.

C.L. BORGMAN and J. FURNER (2002). Scholarly communication and bibliometrics. In: B. Cronin (ed.). Annual Review of Information Science and Technology 36. Medford, NJ: Information Today, 3-72.

S. BORNHOLDT and H. EBEL (2001). World Wide Web scaling exponent from Simon's 1955 model. Physical Review E 64(3), 035104R (4 pages).

T. BRAUN, W. GLÄNZEL and A. SCHUBERT (2005). A Hirsch-type index for journals. The Scientist 19(22), 8.

T. BRAUN, W. GLÄNZEL and A. SCHUBERT (2006). A Hirsch-type index for journals. Scientometrics 69(1), 169-173.

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

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

L. EGGHE (2008). Modelling successive h-indices. Scientometrics 77(3), to appear.

L. EGGHE, M. GOOVAERTS and H. KRETSCHMER (2008). Collaboration and productivity: an investigation in Scientometrics and in a university repository. COLLNET Journal of Scientometrics and Information Management, to appear.

L. EGGHE and I.K. RAVICHANDRA RAO (2008). Study of different h-indices for groups of authors. Journal of the American Society for Information Science and Technology 59(6), 1-6.

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

L. EGGHE and R. ROUSSEAU (2008). An h-index weighted by citation impact. Information Processing and Management, to appear.

M. FALOUTSOS, P. FALOUTSOS and C. FALOUTSOS (1999). On power-law relationships of the Internet topology. Computer Communication Review 29, 251-263.

G. HERDAN (1960). Type-Token Mathematics. A Textbook of mathematical Linguistics. Mouton, 's Gravenhage, the Netherlands.

J.E. HIRSCH (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.

B.A. HUBERMAN (2001). The Laws of the Web. Patterns in the Ecology of Information. The MIT Press, Cambridge (MA), USA.

B.A. HUBERMAN, P.L.T. PIROLLI, J.E. PITKOW and R.M. LUKOSE (1998). Strong regularities in World Wide Web surfing. Science 280, 95-97.

Y.M. IOANNIDES and H.G. OVERMAN (2003). Zipf's law for cities: an empirical examination. Regional Science and Urban Economics 33, 127-137.

H. JEONG, B. TOMBOR, R. ALBERT, Z.N. OTTVAL and A.-L. BARABÁSI (2000). The large-scale organization of metabolic networks. Nature 407, 651-654.

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.

P.J. LAMBERT (2001). The Distribution and Redistribution of Income. Third Edition. Manchester University Press, Manchester.

A.J. LOTKA (1926). The frequency distribution of scientific productivity. Journal of the Washington Academy of Sciences 16(12), 317-324.

M. MARSILI and Y.-C. ZHANG (1998). Interacting individuals leading to Zipf's law. Physical Review Letters 80(12), 2741-2744.

J. NIELSEN (1997). Do websites have increasing returns? http://www.useit.com/alertbox/9704b.html

M.L. PAO (1981). Co-authorship as communication measure. Library Research 2, 327-338.

M.L. PAO (1982). Collaboration in computational musicology. Journal of the American Society for Information Science 33(1), 38-43.

M.L. PAO (1992). Global and local collaborators: a study of scientific collaboration. Information Processing and Management 28(1), 99-109.

R. PASTOR-SATORRAS and A. VESPIGNANI (2004). Evolution and Structure of the Internet. A statistical Physics Approach. Cambridge University Press, Cambridge, UK.

G. PRATHAP (2006). Hirsch-type indices for ranking institution's scientific research output. Current Science 91(11), 10 December 2006, 1439.

D.J. DE SOLLA PRICE and D. DEB. BEAVER (1966). Collaboration in an invisible college. American Psychologist 21, 1011-1018.

R. ROUSSEAU (1997). Sitations: an exploratory study. Cybermetrics 1(1), paper 1, 1997. http://www.cindoc.csis.es/cybermetrics/articles/v1i1p1.html

M. SCHREIBER (2007). The influence of self-citation corrections on Egghe's g-index. Scientometrics 76(1), 2008, to appear.

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

K. SUBRAMANYAM (1983). Bibliometric studies of research collaboration. Journal of Information Science 6, 33-38.

M. THELWALL and D. WILKINSON (2003). Graph structure in three national academic webs: power laws and anomalies. Journal of the American Society for Information Science and Technology 54(8), 706-712.

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

G.K. ZIPF (1949). Human Behavior and the Principle of least Effort. Addison-Wesley, Cambridge, USA. Reprinted: Hafner, New York, USA, 1965.

H. ZUCKERMAN (1967). Nobel laureates in science: patterns of productivity, collaboration and authorship. American Sociological Review 32(3), 391-403.