An explanation of disproportionate growth using linear 3-dimensional informetrics and its relation with the fractal dimension

Peer-reviewed author version

EGGHE, Leo (2005) An explanation of disproportionate growth using linear 3-dimensional informetrics and its relation with the fractal dimension. In: SCIENTOMETRICS, 63(2). p. 277-296.

DOI: 10.1007/s11192-005-0213-3
Handle: http://hdl.handle.net/1942/756

An explanation of disproportionate growth using linear 3-dimensional informetrics and its relation with the fractal dimension

by

L. Egghe

Limburgs Universitair Centrum (LUC), Universitaire Campus, B-3590 Diepenbeek, Belgium[1]
and
Universiteit Antwerpen (UA), Campus Drie Eiken, Universiteitsplein 1, B-2610 Wilrijk, Belgium

leo.egghe@luc.ac.be

ABSTRACT

We study new and existing data sets which show that growth rates of sources usually are different from growth rates of items. Examples: references in publications grow with a rate that is different (usually higher) from the growth rate of the publications themselves; article growth rates are different from journal growth rates and so on. In this paper we interpret this phenomenon of "disproportionate growth" in terms of Naranan's growth model and in terms of the self-similar fractal dimension of such an information system, which follows from Naranan's growth model.

---

[1] Permanent address

Acknowledgements: The author is grateful to prof. Dr. R. Rousseau for interesting discussions on the topic of this paper and to M. Sc. M. Goovaerts for constructing the dataset from Econlit, which we use in this paper.

The main part of the paper is devoted to explain disproportionate growth. We show that the "simple" 2-dimensional informetrics models of source-item relations are not able to explain this but we also show that linear 3-dimensional informetrics (i.e. adding a new source set) is capable to model disproportionate growth. Formulae of such different growth rates are presented using Lotkaian informetrics and new and existing data sets are presented and interpreted in terms of the used linear 3-dimensional model.

I. Introduction

In the paper Persson, Glänzel and Danell (2003), the authors present interesting "universal" data (i.e. based on all papers indexed in the volumes 1980-2000 of the Science Citation Index® (SCI)). One of the data they mention is that, between 1980 and 1998, the number of articles has increased by (roughly) one third whereas the number of citations received by them has increased by three quarters. In other words, denoting by $n(t)$ (t = 1980 or 1998) the number of articles at time (year) t and by $c(t)$ the number of citations at time t we have that

$$\frac{n(1980)}{n(1998)} = \frac{4}{3} \qquad (1)$$

but

$$\frac{c(1980)}{c(1998)} = \frac{7}{4} \qquad (2)$$

In more general notation we can state that for $t_1 < t_2$ we have that

$$\frac{n(t_2)}{n(t_1)} \neq \frac{c(t_2)}{c(t_1)} \qquad (3)$$

where, in the above example, we can replace $\neq$ by $<$.

How to explain this "disproportionate growth" as this phenomenon is (rightly) called in Persson, Glänzel and Danell (2003)? Also correct is their statement that if we would expect reference lists to become longer simply because there are more articles to cite, then we can only conclude an equality in (3), i.e. proportionate growth, hence no explanation for disproportionate growth. They further indicate that the only way to understand the phenomena described above is to consider – in addition to articles and citations – author activity, where activity can be expressed in different forms: papers per author, author collaboration, etc. This idea will be further developed in this article by considering linear 3-dimensional informetrics of the type discussed in Egghe (2003, 2004a), i.e. informetrics of the form (in the terminology of above):

$$\text{authors} \rightarrow \text{articles} \rightarrow \text{citations}$$

where "$\rightarrow$" means "produce", i.e. the informetrics of positive reinforcement where we have authors writing articles and where articles have (or receive) citations.

Before producing more (known and new) data sets, illustrating disproportionate growth, let us first make a remark on yearly growth (or growth per unit time period) versus cumulative growth. The numbers $n(t)$ and $c(t)$ above indicate, respectively, the number of articles with publication year t and the number of citations received (or given) in year t, of course all limited e.g. to a fixed field. We can, of course, also consider all articles published at t or before t and the same goes for the citations. In mathematical terminology and using time t as a continuous variable we can hence define cumulative growth of articles and citations as follows. Let $N(t)$ denote the cumulative number of articles published up to time t (we suppose t = 0 to be early enough that no publications appeared before t = 0). It is clear that

$$N(t) = \int_0^t n(t')dt' \tag{4}$$

for every $t \geq 0$. In the same way we can define

$$C(t) = \int_0^t c(t')dt' \quad (5)$$

as the cumulative number of citations up to time t.

It is clear that disproportionate growth, as expressed by (3) also leads to disproportionate growth when we use the functions N and C instead of n and c, but the proportions are different. Let us indicate, however, the following easy proposition.

Proposition I.1 :
Let the functions n and c be represented as increasing exponential functions, i.e. of the type

$$h(t) = ca^t \quad (6)$$

where $t \geq 0$, $a > 1$, $c > 0$ (c, a: parameters)
Then, in the limit for $t_1 \to \infty$, growth rates

$$\frac{h(t_2)}{h(t_1)} \quad (7)$$

(for $t_1 < t_2$), as in (3), are the same as cumulative growth rates

$$\frac{H(t_2)}{H(t_1)} \quad (8)$$

where

$$H(t) = \int_0^t h(t')dt' \quad (9)$$

as in (4) or (5).

Proof : We have

$$H(t) = \int_0^t h(t')dt'$$

$$H(t) = \int_0^t c a^{t'} dt'$$

$$H(t) = \frac{c}{\ln a}\left(a^t - 1\right)$$

Hence, for $t_1 < t_2$,

$$\frac{H(t_2)}{H(t_1)} = \frac{a^{t_2} - 1}{a^{t_1} - 1}$$

So

$$\frac{\frac{H(t_2)}{H(t_1)}}{\frac{h(t_2)}{h(t_1)}} = \frac{\frac{a^{t_2} - 1}{a^{t_1} - 1}}{a^{t_2 - t_1}}$$

$$= \frac{a^{t_1} - a^{t_1 - t_2}}{a^{t_1} - 1}$$

$$= \frac{1 - a^{-t_2}}{1 - a^{-t_1}}$$

which goes to 1 for t_1 (hence $t_2 > t_1$) going to ∞ (since $a > 1$). ~

In words, for t_1 large, i.e. far away from the starting point of the database under consideration and for any $t_2 > t_1$, we have, by applying Proposition I.1 to $h = n$ and $h = c$ (hence $H = N$ and $H = C$ respectively)

$$\frac{n(t_2)}{n(t_1)} \approx \frac{N(t_2)}{N(t_1)} \tag{10}$$

and

$$\frac{c(t_2)}{c(t_1)} \approx \frac{C(t_2)}{C(t_1)} \tag{11}$$

showing that disproportionate (e.g.) yearly growth (as in (3)) implies disproportionate cumulative growth and the "degrees of disproportionality" are the same. Note that this is only a limiting result. In the other cases we will have clear differences between both growth types.

In the sequel we will study both cases of growth: some results will be valid for both cases and some only for cumulative growth rates.

In the next section we formally define "disproportionate growth" and indicate its relation with Naranan's growth model (Naranan (1970), Egghe (2004b)). These results are only valid for cumulative growth data. We also indicate how one can calculate the fractal dimension of the system, given cumulative growth data, based on Egghe (2004b). New and existing examples are thus analyzed.

In the third section we give a possible explanation of disproportionate growth, valid for cumulative and non-cumulative growth. This is done by using a new source set as will become clear in the following examples

1. In the case of growth of articles and references (or citations) we also consider authors who publish these articles. Alternatively, one can also consider journals that publish these articles.
2. In the case of growth of journals and articles, we can also consider publishers who produce these journals.

In each of these cases, to the classical informetric production process (IPP) in 2 dimensions, i.e.

sources ® items

(sources producing items) we add a new source set, i.e.

new source set ® sources ® items

The new sources produce the old sources which in turn produce the items (e.g. publishers produce journals and these journals produce articles). We hence have the framework of linear 3-dimensional informetrics as developed in Egghe (2003, 2004a). We show that disproportionate growth of items with respect to the original sources is covered by linear 3-dimensional informetrics (also called positive reinforcement) and is <u>not</u> covered by 2-dimensional informetrics. Formulae are presented, yielding disproportionate growth rates in terms of the parameters appearing in linear 3-dimensional informetrics (we also briefly discuss the results of this theory, for the sake of completeness).

II. Disproportionate growth and its relation to Naranan's growth model and to the theory of self-similar fractals

Let us consider an IPP, say at time t_1, where there are T_1 sources and A_1 items. Let us then take a second IPP, say at time $t_2 > t_1$, where there are T_2 sources and A_2 items. This second IPP can be considered as "grown" out of the first one from t_1 to t_2. In this general setting we can consider cumulative or non-cumulative numbers.

Definition II.1:

Define

$$\theta = \frac{T_2}{T_1} \tag{12}$$

and

$$\eta = \frac{A_2}{A_1} \tag{13}$$

1. We say that we have disproportionate growth in the positive sense if

$$\eta > \theta \tag{14}$$

2. We say that we have disproportionate growth in the negative sense if

$$\eta < \theta \tag{15}$$

3. We say that we have proportionate growth if

$$\eta = \theta \tag{16}$$

Of course, when using practical data, we will conclude that we have (approximately) proportional growth if $\eta \approx \theta$, even if $\eta \neq \theta$.

Examples II.2

1. The following data set has been compiled by M. Goovaerts (LUC) on the topic “management” in Econlit, where we restricted ourselves, naturally, to journal articles. The data concern cumulative growth. The topic “management” was chosen because it yields a large (but still manageable) data set hence a stable growth profile.

Table 1. Cumulative growth in Econlit on the topic "management"

	Articles	Journals
1990	9,545	402
1991	10,688	448
1992	12,047	499
1993	13,534	534
1994	15,183	576
1995	17,347	627
1996	19,788	693
1997	22,262	748
1998	25,098	794
1999	28,175	837
2000	31,334	880
2001	34,636	915
2002	37,889	946

If we compare the first line (1990) with the last one (2002) we see that $\theta = 2.35$ while $\eta = 3.97$, a clear disproportionate growth in the positive sense. An explanation will be given in the next section.

2. The following data set can be found in Wu et al. (2003) on scores of Chinese papers in the SCI (Science Citation Index). Here not only the growth of papers and citations received are presented but also the number of cited papers. In all cases the numbers refer to non-cumulative growth

Table 2. Non-cumulative growth of Chinese SCI papers and their citations

	Articles	Citations	Cited Articles
1991	6,630	6,771	3,608
1992	6,224	11,384	5,994
1993	9,617	12,896	7,060
1994	10,411	12,626	7,180
1995	13,134	14,000	7,869
1996	14,459	15,800	8,826
1997	16,883	18,434	9,952
1998	19,838	21,511	11,549
1999	24,476	25,173	13,024
2000	30,499	31,384	15,733

The comparison of the years 1991 and 2000 for Articles-Citations yields $\theta = 4.60 \approx \eta_1 = 4.64$ while the comparison Articles-Cited Articles (such a comparison is not made very often in the literature) we have $\theta = 4.60 \approx \eta_2 = 4.36$. We can conclude that Articles-Citations shows a proportionate growth and the same can be said about the comparison Articles-Cited Articles although there is a slight disproportionate growth in the negative sense $(\eta_2 = 4.36 < 4.60 = \theta)$.

However, in the same article, data on CSTPC papers (i.e. papers in the China Scientific and Technical Papers and Citation database) and their number of references shows a clear disproportionate growth in the positive sense (non-cumulative data).

Table 3. Non-cumulative growth of CSTPC papers and their references

	Papers	References
1994	107,492	100,748
1995	107,991	104,758
1996	116,239	172,385
1997	120,851	280,476
1998	133,341	335,314
1999	162,779	466,611
2000	180,848	554,324

When we compare the first line (1994) with the last line (2000) we see that $\theta = 1.68$ and $\eta = 5.50$, hence a clear disproportionate growth in the positive sense. Note that the data in Tables 2 and 3 can also be converted into cumulative data, yielding comparable conclusions.

We now describe Naranan's growth model of sources and items, cf. Naranan (1970), Egghe (2004b). In the evolution of an IPP over time t, there are two growth processes in Naranan's model: one of number of sources and one of number of items in each source. Both are considered to be exponential growth models, the simplest but most basic growth model.

(i) The number of sources, $N(t)$, at time t grows exponentially:

$$N(t) = c_1 a_1^t \tag{17}$$

$c_1 > 0$, $a_1 > 1$,

(ii) The number of items, $p(t)$, in each source grows exponentially with the same parameters for every source:

$$p(t) = c_2 a_2^t \tag{18}$$

$c_2 > 0$, $a_2 > 1$.

We have the following easy theorem (extending Naranan's model above to all values $a_2 > 0$):

Theorem II.3:
Adopting Naranan's model, we have, in the notation of Definition II.1, for cumulative growth, and taking the time span between the two IPPs to be 1 (i.e. $t_2 - t_1 = 1$) that

$$a_1 = \theta \tag{19}$$

$$a_2 = \frac{\eta}{\theta} \tag{20}$$

and hence we have disproportionate growth in the positive (resp. negative) sense iff $a_2 > 1$ (resp. $0 < a_2 < 1$) and we have proportionate growth iff $a_2 = 1$ (Note that only the second growth process is needed to characterise (dis)proportionate growth).

Proof:
By the very definition of a_1 and a_2 we have (since $t_2 - t_1 = 1$ and since we consider cumulative growth) that

$$\frac{N(t_2)}{N(t_1)} = a_1 = \frac{T_2}{T_1} = \theta$$

and

$$\frac{p(t_2)}{p(t_1)} = a_2 = \frac{\frac{A_2}{T_2}}{\frac{A_1}{T_1}}$$

$$a_2 = \frac{\eta}{\theta}.$$

The rest follows from Definition II.1. ~

As proved in Egghe (2004b) the model of Naranan also describes such an IPP as a self-similar fractal (see Mandelbrot (1977), Falconer (1985) or Feder (1988) for a description of self-similar fractals). Its fractal dimension D_S (i.e. its fractal complexity) is given by the formula (cf. Egghe (2004b, 2005)), valid for $a_1, a_2 > 1$:

$$D_S = \frac{\ln a_1}{\ln a_2} \tag{21}$$

Formula (21) is obtained by considering an IPP, as described in (17) and (18), as a self-similar fractal. It is then well-known (see Feder (1988), Falconer (1985)) that (21) yields the fractal dimension of such a self-similar fractal. The link with self-similar fractals is explained, in full detail, in Egghe (2004b, 2005).

Consequently, in terms of disproportionate growth in the positive sense, we have the following theorem.

<u>Theorem II.4</u>:

In the notation of Definition II.1 we have that the fractal dimension D_S of such a growing system (where $t_2 - t_1 = 1$) is given by

$$D_S = \frac{\ln\theta}{\ln\eta - \ln\theta} \qquad (22)$$

where we consider cumulative disproportionate growth in the positive sense.

<u>Proof</u>:

This follows readily from Theorem II.3, (21), the fact that $a_2 > 1$ (i.e. $D_S > 0$). ~

In the above, we have supposed (for the interpretation of Naranan's model in terms of (dis)proportionate growth) that the time period between the two systems equals 1: $t_2 - t_1 = 1$. For practical purposes, to derive the fractal dimension directly from practical data, involving more than one year, we need a formula for D_S in these (more general) cases. The next theorem shows that formula (22) can also be used for other time periods $t_2 - t_1 \neq 1$.

<u>Theorem II.5</u>:

Suppose we have a cumulative growth system as in Definition II.1 with $t_2 - t_1 \neq 1$ (but $t_2 - t_1 > 0$). Then formula (22) is still valid for the fractal dimension of such a system in case of disproportionate growth in the positive sense.

<u>Proof</u>:

Considering the time period $t_2 - t_1$ as unit time requires the transformation

$$t \rightarrow \frac{t}{t_2 - t_1} = t' \qquad (23)$$

In Naranan's model this has, as a consequence that

$$N\left(t^{'}\right) = c_1\left(a_1^{t_2 - t_1}\right)^{t^{'}} \tag{24}$$

$$p\left(t^{'}\right) = c_2\left(a_2^{t_2 - t_1}\right)^{t^{'}} \tag{25}$$

Hence

$$N\left(t^{'}\right) = c_1 a_1^{'t^{'}} \tag{26}$$

$$p\left(t^{'}\right) = c_2 a_2^{'t^{'}} \tag{27}$$

where $a_1^{'} = a_1^{t_2 - t_1} > 1$ and $a_2^{'} = a_2^{t_2 - t_1} > 1$ and where the time span $t_2^{'} - t_1^{'}$ between the two IPPs is $\frac{t_2}{t_2 - t_1} - \frac{t_1}{t_2 - t_1} = 1$. According to Definition II.1 and Theorem II.3 we have that

$$a_1^{'} = \theta \tag{28}$$

$$a_2^{'} = \frac{\eta}{\theta} \tag{29}$$

and according to (21) we have that the fractal dimension of this system is (since $a_1^{'}, a_2^{'} > 1$)

$$D_S^{'} = \frac{\ln a_1^{'}}{\ln a_2^{'}}. \tag{30}$$

Hence, formula (22) for $t_2 - t_1 \neq 1$ follows from (28), (29) and (30). ~

Another way to describe that the length of the time period is not involved in the calculation of the fractal dimension, as long as the same disproportionate growth is occurring is as follows. Let, in the first case, we have parameters θ and η as in Definition II.1 with $\eta > \theta$. Suppose we extend (or shorten) the growing period by a factor $\gamma \neq 1, \ \gamma > 0$. It is clear that the new growth parameters of the second system are (since the rates remain the same)

$$\theta' = \theta^{\gamma} \tag{31}$$

$$\eta' = \eta^{\gamma} \tag{32}$$

and that $\eta' > \theta'$. By the above theorem we have that the fractal dimension of the second system is

$$D_S' = \frac{\ln\left(\theta^{\gamma}\right)}{\ln\left(\eta^{\gamma}\right) - \ln\left(\theta^{\gamma}\right)}$$

$$D_S' = \frac{\ln\theta}{\ln\eta - \ln\theta} = D_S, \tag{33}$$

i.e. the fractal dimension of the original system. This also follows if we adapt the proof of Theorem II.3 to time periods $t_2 - t_1 \neq 1$.

<u>Note II.6.</u>

The interpretation of disproportionate growth in terms of Naranan's model could only be given for cumulative growth, by condition (i) in this model (N(t) denotes the total number of sources at time t). Because of (ii) in this model (i.e. $a_2 > 1$) we only have an interpretation of disproportionate growth in the positive sense. Note that this is the most important case: most examples show a faster growth of the items than the one of the sources.

In case of cumulative disproportionate growth in the negative sense, we can reverse the roles of sources and items (and replace the verb "produce" by "is produced by"), i.e. the dual interpretation of the original IPP. For this dual IPP, the results above are valid since we now have disproportionate growth in the positive sense.

We close this section with an example of the calculation of the fractal dimension.

Example II.7.

We use the cumulative growth data of Table 1. In this table, each couple of 2 years can be used to calculate the fractal dimension. The most stable result is obtained when we use the first and the last line. We calculated already that $\theta = 2.4$ and $\eta = 3.97$. By theorems II.4 and II.5 we have that the fractal dimension is

$$D_S = \frac{\ln(2.35)}{\ln\left(\frac{3.97}{2.35}\right)} = 1.6295$$

Note (Egghe and Rousseau (1990), Egghe (2004 b, 2005)) that this theory is linked with Lotkaian informetrics, i.e. where the law of Lotka

$$f(j) = \frac{C}{j^\alpha} \tag{34}$$

is valid as size-frequency function. We have that

$$\alpha = 1 + D_S \tag{35}$$

hence, for the above example we find the value $\alpha = 2.6295$ for Lotka's exponent.

The next section is devoted to explaining disproportionate growth.

III. Explanation of disproportionate growth using linear 3-dimensional informetrics

III.1. Introduction

The previous section linked disproportionate growth with the value of the second growth parameter a_2 in Naranan's model. The value of this approach lies in the link of disproportionate growth with fractal theory and Lotkaian informetrics. The theory developed in the previous section does not, however, explain the occurrence of disproportionate growth: disproportionate growth simply is equivalent – in Naranan's model – with certain values of the parameter a_2 but, trying to consider this as an explanation leaves us with the explanation of the a_2-values themselves.

Since, in the definition of disproportionate growth (Definition II.1), we only deal with sources and items (so-called 2-dimensional informetrics) one is inclined to think that an explanation of disproportionate growth follows from such a 2-dimensional theory. This is, however, not true.

Indeed, just to keep things as simple as possible, let us repeat some elementary results from 2-dimensional Lotkaian informetrics, i.e. a system where (34) is valid and let us suppose $\alpha > 2$ and $j \in [1,+\infty[$, the range of possible item per source densities (see Egghe and Rousseau (1990), Egghe (1989, 1990), Egghe (2005)). Then we have that

$$T = \int_1^{\infty} f(j)dj = \frac{C}{\alpha - 1} \tag{36}$$

and

$$A = \int_1^{\infty} jf(j)dj = \frac{C}{\alpha - 2} \tag{37}$$

is the total number of sources (T) and the total number of items (A) in this system. Hence we have

$$\frac{A}{T} = \frac{\alpha - 1}{\alpha - 2} \tag{38}$$

Suppose now this system grows from T to T' (number of sources) and from A to A' (number of items). Since we keep the subject fixed it is most natural to assume that, in the new system, we have (34) with the same α. Hence from (38) we have

$$\frac{A'}{T'} = \frac{\alpha - 1}{\alpha - 2} = \frac{A}{T}$$

hence (using the notation of Definition II.1)

$$\eta = \frac{A'}{A} = \frac{T'}{T} = \theta. \tag{39}$$

Consequently, 2-dimensional informetrics, keeping α fixed, can only explain proportionate growth. But even if α changes, these changes need to be explained but there is no tool available (to the best of my knowledge) in 2-dimensional informetrics to accomplish this.

The argument leading to formula (39) is an exact interpretation of the fact that growth of items, as a consequence of growth of sources only (as e.g. described in Persson, Glänzel and Danell (2003) in the case of articles and references), can only be proportionate. As indicated in Persson, Glänzel and Danell (2003), only the introduction of a "third party" (in their case "authors") can explain disproportionate growth. This brings us to 3-dimensional informetrics which we will briefly describe in Subsection III.3 (for more details, we refer the reader to Egghe (2003, 2004a). In Subsection III.4 we will then show that this theory is capable of explaining the different growth rates η and θ of Definition II.1, hence of explaining disproportionate growth.

III.2. Elements of linear 3-dimensional informetrics

Suppose we have a 2-dimensional IPP where there are T sources and A items. Now we consider the T sources as being produced by S new sources (i.e. "super sources").

Examples:

1. Authors (super sources) write papers (sources) and these papers have references (items) (but, of course, given by these authors, i.e. these super sources).
2. Publishers (super sources) produce journals (sources) and these journals contain articles (items) (being added to the journal volumes by these publishers, i.e. these super sources).

These examples make it clear that the super sources are not only involved in producing the sources but also in the linking of the items to these sources. This remark will turn out to be basic for the explanation of disproportionate growth of sources and items.

The formal description is as follows (continuous setting), see Egghe (2003, 2004a). Let our original source set be given by the interval $[0,T]$ and the original item set be given by $[0,A]$ $(A > T > 0)$. The "addition" of a new "super source" set $[0,S]$ leads to the following two IPPs. First of all we have the one describing the 2-dimensional informetric theory between $[0,S]$ and $[0,T]$. Let us suppose that we have here a Lotkaian size-frequency function

$$f(j) = \frac{B}{j^{\alpha}} \tag{40}$$

with $\alpha > 2$ (for the sake of simplicity; extensions to $0 < \alpha \leq 2$ are possible). Next we also have the IPP describing the 2-dimensional informetric theory between $[0,S]$ and $[0,A]$. It is clear that here the size-frequency function, denoted f, is dependent on α <u>and</u> on how the super sources determine the items in the sources (i.e. a composite action, i.e. 3-dimensional informetrics). In the Lotkaian model it is shown in Egghe (2003, 2004a) that

$$f^*(j) = \frac{C}{j^{\frac{\alpha+\beta-1}{\beta}}} \tag{41}$$

where $\beta > 0$ depends on this composite action (the α here is the α_1 in Egghe (2003, 2004a)).

III.3. The modelling of disproportionate growth

Let us now consider two of such models (indexed with 1 and 2). Schematically, we have a situation as in Fig. 1.

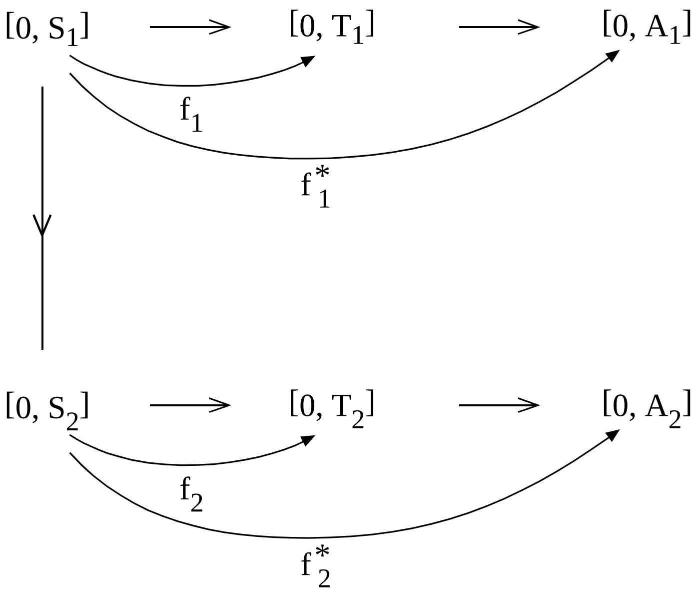

Fig. 1. Growth in 3 dimensions

The second system is considered as a grown version of the first system (we do not specify how the growth process took place nor we need a time span specification between the two systems).

Let us suppose that the size-frequency function f_1 satisfies (cf. (40)), for $j \geq 1$

$$f_1(j) = \frac{B_1}{j^{\alpha_1}} \tag{42}$$

and, according to (41), let the size-frequency function f_1^* be given by (for $j \geq 1$)

$$f_1^*(j) = \frac{C_1}{j^{\frac{\alpha_1+\beta_1-1}{\beta_1}}} \tag{43}$$

Since the second system has grown from the first one we can assume that f_2 is a Lotka function with the same exponent α_1, since we are still considering the same informational topic. Hence we have for $j \geq 1$

$$f_2(j) = \frac{B_2}{j^{\alpha_1}} \tag{44}$$

According to (41) we now have a second parameter β_2 such that f_2^* satisfies, for $j \geq 1$

$$f_2^*(j) = \frac{C_2}{j^{\frac{\alpha_1+\beta_2-1}{\beta_2}}} \tag{45}$$

We further assume all exponents of j in formulae (42)-(45) to be superior to 2 (note that this requires α_1 to be superior to max(2, $1+\beta_1$, $1+\beta_2$)). We have the following Theorem.

Theorem III.1:
Under the above assumptions we have that

$$\frac{A_2}{A_1} = \frac{\alpha_1 - \beta_1 - 1}{\alpha_1 - \beta_2 - 1} . \frac{T_2}{T_1} . \tag{46}$$

We hence have disproportionate growth in the positive sense iff $\beta_1 < \beta_2$, disproportionate growth in the negative sense iff $\beta_1 > \beta_2$ and proportionate growth (also included in this model!) iff $\beta_1 = \beta_2$.

Proof:

Since all exponents of j in formulae (42)-(45) are assumed to be superior to 2 we have, by the very definition of f_1, f_1^*, f_2 and f_2^* that, using repeatedly (38) for the appropriate values of A and T, applicable to f_1, f_1^*, f_2, f_2^*:

$$\frac{T_1}{S_1} = \frac{\int_1^{\infty} j f_1(j)\,dj}{\int_1^{\infty} f_1(j)\,dj} = \frac{\alpha_1 - 1}{\alpha_1 - 2} \tag{47}$$

$$\frac{T_2}{S_2} = \frac{\int_1^{\infty} j f_2(j)\,dj}{\int_1^{\infty} f_2(j)\,dj} = \frac{\alpha_1 - 1}{\alpha_1 - 2} \tag{48}$$

$$\frac{A_1}{S_1} = \frac{\int_1^{\infty} j f_1^*(j)\,dj}{\int_1^{\infty} f_1^*(j)\,dj} = \frac{\frac{\alpha_1 + \beta_1 - 1}{\beta_1} - 1}{\frac{\alpha_1 + \beta_1 - 1}{\beta_1} - 2} = \frac{\alpha_1 - 1}{\alpha_1 - 1 - \beta_1} \tag{49}$$

$$\frac{A_2}{S_2} = \frac{\int_1^{\infty} j f_2^*(j)\,dj}{\int_1^{\infty} f_2^*(j)\,dj} = \frac{\frac{\alpha_1 + \beta_2 - 1}{\beta_2} - 1}{\frac{\alpha_1 + \beta_2 - 1}{\beta_2} - 2} = \frac{\alpha_1 - 1}{\alpha_1 - 1 - \beta_2} \tag{50}$$

Formulae (47) and (48) yield

$$\frac{T_1}{S_1} = \frac{T_2}{S_2}$$

hence

$$\frac{S_2}{S_1} = \frac{T_2}{T_1} \tag{51}$$

(in fact, we refind the result of Subsection III.1). Formulae (49) and (50) yield

$$\frac{\frac{A_2}{S_2}}{\frac{A_1}{S_1}} = \frac{\alpha_1 - 1 - \beta_1}{\alpha_1 - 1 - \beta_2}$$

, so

$$\frac{A_2}{A_1} = \frac{\alpha_1 - 1 - \beta_1}{\alpha_1 - 1 - \beta_2} . \frac{S_2}{S_1} \quad (52)$$

Combining (51) and (52) yields formula (46). The rest of the assertions are trivial from this.

~

Corollary III.2:

In terms of Naranan's growth model we find that

$$a_2 = \left(\frac{\alpha_1 - \beta_1 - 1}{\alpha_1 - \beta_2 - 1}\right)^{\frac{1}{t}} \quad (53)$$

where t is the time span between both IPPs.

Proof:

Since

$$p(t) = c_2 a_2^t$$

we have that

$$\frac{p(t)}{p(0)} = a_2^t \quad (54)$$

Taking 0 as the time of the first IPP (and hence t as the time of the second IPP) we have that, by definition of p,

$$\frac{p(t)}{p(0)} = \frac{\frac{A_2}{T_2}}{\frac{A_1}{T_1}}$$

$$\frac{p(t)}{p(0)} = \frac{\frac{A_2}{A_1}}{\frac{T_2}{T_1}}$$

$$\frac{p(t)}{p(0)} = \frac{\alpha_1 - \beta_1 - 1}{\alpha_1 - \beta_2 - 1} \qquad (55)$$

by (46). Hence (53) follows from (54) and (55). ~

<u>Note III.2</u>:

Suppose we have that $S_1 < S_2$ in the above model. From this, based on (43) and (45), there is no general relation between β_1 and β_2 possible. Indeed, as follows from Egghe (2004c) (see also Egghe (2005), Chapter II), we have that $f(j) = \frac{1,000}{j^3}$ is the size-frequency function, given 500 sources and 1,000 items while for 1,000 sources and 10,000 items we have the Lotka function $f(j) = \frac{1,111}{j^{2.11}}$, while for 2,000 sources and 4,000 items we have the function $f(j) = \frac{4,000}{j^3}$. Interpreted for super sources and items, it follows from the above and from (43) and (45), that $S_1 < S_2$ can imply $\beta_1 < \beta_2$ but also $\beta_1 > \beta_2$.

IV. Conclusions

In this paper we defined disproportionate growth (of items versus sources) in the positive sense (item growth is relatively faster than source growth) or in the negative sense (opposite effect) or proportionate growth (both growth rates are the same). We showed the relation of these notions with Naranan's growth model (Naranan (1970), see also Egghe (2004b)) and showed how one can calculate, from practical cumulative growth data, the fractal dimension of the system (independent of the time span).

The paper further shows that an explanation of disproportionate growth of sources and items cannot be obtained from 2-dimensional informetrics involving these sources and items. This surprising result and a suggestion in Persson, Glänzel and Danell (2003) made us think of applying the 3-dimensional informetrics theory as developed in Egghe (2003, 2004a). The explanation consists of "defining" a set of "super sources" which produce the original sources but which also attach the items into the original sources. In this way, disproportionate growth of references versus articles can be explained by looking at authors. In the same way, disproportionate growth of articles versus journals (a new dataset is compiled from the database Econlit) can be explained by considering journal publishers.

Applying the mathematical theory of Lotkaian 3-dimensional informetrics, as developed in Egghe (2003, 2004a), yields a formula of the growth rate of items in function of the growth rate of sources, involving a parameter that is determined by the growth of super sources. This parameter explains both types of disproportionate growth and even proportionate growth.

References

L. Egghe (1989). The Duality of informetric Systems with Applications to the empirical Laws. Ph. D. Thesis, City University, London (UK), 1989.

L. Egghe (1990). The duality of informetric systems with applications to the empirical laws. Journal of Information Science 16(1), 17-27, 1990.

L. Egghe (2003). Positive reinforcement and 3-dimensional informetrics. Proceedings of the ninth International Conference on Scientometrics and Informetrics, Beijing (China), Jiang Guohua, R. Rousseau and Wu Yishan, eds., 47-54, Dalian University of Technology Press, Dalian (China), 2003.

L. Egghe (2004a). Positive reinforcement and 3-dimensional informetrics. Scientometrics 60(3), 497-509, 2004.

L. Egghe (2004b). The power of power laws and the interpretation of Lotkaian informetric systems as self-similar fractals. Journal of the American Society for Information Science and Technology, to appear, 2004.

L. Egghe (2004c). The source-item coverage of the Lotka function. Scientometrics 61(1), 103-115, 2004.

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

L. Egghe and R. Rousseau (1990). Introduction to Informetrics. Quantitative Methods in Library, Documentation and Information Science. Elsevier, Amsterdam, 1990.

K.J. Falconer (1985). The Geometry of fractal Sets. Cambridge Tracts in Mathematics 85. Cambridge University Press, Cambridge (UK), 1985.

J. Feder (1988). Fractals. Plenum, New York, 1988.

B. Mandelbrot (1977). The fractal Geometry of Nature. Freeman, New York, 1977.

S. Naranan (1970). Bradford’s law of bibliography of science: an interpretation. Nature 227(5258), 631-632, August 8, 1970.

O. Persson, W. Glänzel and R. Danell (2003). Inflationary bibliometric values: the role of scientific collaboration and the need for relative indicators in evaluative studies. Proceedings of the ninth International Conference on Scientometrics and Informetrics, Beijing (China), Jiang Guohua, R. Rousseau and Wu Yishan, eds., 411-420, Dalian University of Technology Press, Dalian (China), 2003.

Yishan Wu, Yuntao Pan, Yuhua Zhang, Zheng Ma, Jingan Pang, Hong Guo, Bo Xu and Zhiqing Yang (2003). China Scientific and Technical Papers and Citations (CSTPC): history, impact and outlook. Proceedings of the ninth International Conference on Scientometrics and Informetrics, Beijing (China), Jiang Guohua, R. Rousseau and Wu Yishan, eds., 352-361, Dalian University of Technology Press, Dalian (China), 2003.