**Made available by Hasselt University Library in https://documentserver.uhasselt.be**

General evolutionary theory of information production processes and applications to the evolution of networks

Peer-reviewed author version

EGGHE, Leo (2007) General evolutionary theory of information production processes and applications to the evolution of networks. In: JOURNAL OF INFORMETRICS, 1(2). p. 115-122.

DOI: 10.1016/j.joi.2006.10.003
Handle: http://hdl.handle.net/1942/1783

# General evolutionary theory of information production processes and applications to the evolution of networks

by

L. Egghe

Universiteit Hasselt (UHasselt), Campus Diepenbeek, Agoralaan, B-3590 Diepenbeek, Belgium[1]
and
Universiteit Antwerpen (UA), Campus Drie Eiken, Universiteitsplein 1, B-2610 Wilrijk, Belgium

leo.egghe@uhasselt.be

---

## ABSTRACT

Evolution of information production processes (IPPs) can be described by a general transformation function for the sources and for the items. It generalises the Fellman-Jakobsson transformation which only works on the items.

---

[1] Permanent address
Key words and phrases: evolution, IPP, information production process, Lotka, Zipf, network.
Acknowledgements. The author is grateful to prof. Dr. V. Cothey for mentioning the problem of modelling network evolutions and to profs. Dr. V. Cothey and R. Rousseau for interesting discussions on the topic of this paper. The author is also grateful to two anonymous referees for giving good suggestions to improve this paper, both in content and in style.

In this paper the dual informetric theory of this double transformation, defined by the rank-frequency function, is described by e.g. determining the new size-frequency function. The special case of power law transformations is studied thereby showing that a Lotkaian system is transformed into another Lotkaian system, described by a new Lotka exponent. We prove that the new exponent is smaller (larger) than the original one if and only if the change in the sources is smaller (larger) than that of the items.

Applications to the study of the evolution of networks are given, including cases of deletion of nodes and/or links but also applications to other fields are given.

# I. Introduction

The informetrics of information production processes (IPPs) can be described via the so-called size-frequency function f:

$$\begin{aligned} f : [a, \rho_m] &\to \mathbb{R}^+ \\ j &\to f(j) \end{aligned} \tag{1}$$

where $f(j)$ denotes the density of the sources in item density j: this is the continuous extension for the classical $f(j)$ = number of sources with j items and we let $j \geq a \geq 1$ also be limited to a maximal item density $\rho_m$ (see also Egghe (2005) but there $a = 1$; here we use a general $a > 0$ since we have an application of this case – see further). A classical example is the law of Lotka, where f is then a decreasing power law; this case will also be considered after the general theory.

The size-frequency function f is equivalent with the rank-frequency function g:

$$\begin{aligned} g : [0, T] &\to \mathbb{R}^+ \\ r &\to g(r) \end{aligned} \tag{2}$$

where f and g are related as

$$r = g^{-1}(j) = \int_j^{\rho_m} f(k)dk \tag{3}$$

where $g^{-1}$ denotes the inverse function of g. It is clear from (3) that $g(r)$ denotes the item density in the source on rank density r: this is the continuous extension of the discrete rank-frequency function where $g(r)$ = number of items in the source on rank r and where T denotes (also in the continuous setting) the total number of sources.

The equivalence of the functions f and g is seen as follows: (3) yields $g^{-1}$ (hence g), given f and it follows from (3) that

$$f(j) = -\frac{1}{g'\left(g^{-1}(j)\right)} \tag{4}$$

for all $j \in [a, \rho_m]$, hence f follows from g, showing the equivalency (see also Egghe (2005)). It is also well-known (see Egghe and Rousseau (1990) or Egghe (2005)) that, in case f is a decreasing power law (i.e. Lotka's law), g is the so-called law of Mandelbrot (which we will describe in detail below).

In Egghe (2004), see also Egghe (2003) one studies positive reinforcement of IPPs, where one applies a transformation $\varphi$ on the function g i.e. g is transformed into $g^* = \varphi \circ g$, where $\varphi$ has certain properties, e.g. $\varphi(x) \geq x$ for all x and $\varphi$ strictly increasing. In Egghe (2003, 2004), the connection of positively reinforced IPPs with linear 3-dimensional informetrics (i.e. the composition of 2 IPPs) is highlighted and the concentration properties of these positively reinforced IPPs are indicated using the theorem of Fellman and Jakobsson – see Fellman (1976), Jakobsson (1976); see also Egghe (2006).

In Egghe (2003, 2004) and Egghe and Rousseau (2006a), a transformation of g in the following sense has been studied:

$$g^*(r) = B\left(g(r)\right)^c \tag{5}$$

with B, $c > 1$ (i.e. $\varphi(x) = Bx^c$) yielding, for Lotkaian IPPs, lower Lotka exponents. In Egghe and Rousseau (2006a) the extra generalization $j \in [a, \rho_m]$ with $a \geq 1$ is used. In this case the transformation (5) not only leads to lower Lotka exponents but also to higher minimum density values $a > 1$. This, in turn, gives a rationale for systems in which sources do not have a low number of items as is the case for database sizes or country or city sizes. That these cases go together with low values of the exponent of the Lotka function has been experimentally verified in Egghe and Rousseau (2006a).

Discussions with V. Cothey (July 2005) revealed that an extra generalization of the above formalism (essentially the transformation $g \to \varphi \circ g$) is needed. Indeed, the transformation $\varphi$ is a transformation that applies on the item densities $j = g(r)$ but leaves the source rank densities unchanged. V. Cothey informed us that the framework of IPPs is well applied to networks (where sources are nodes and items are hyperlinks: in- or outlinks) but that a model is needed e.g. to describe disappearing sources (nodes) – of course still allowing for disappearing items as well. Of course the creation of sources and items should also be covered.

In view of the above it is clear what to do: the transformation $\varphi$ above, in its full generality, works well to describe changes (dynamics) of items. So "all we have to do" is to introduce another transformation, called $\psi$ below, in order to describe the changes (dynamics) of the sources.

In the next section the second transformation $\psi$ will act on the rank densities r. So, instead of the transformation

$$g^*(r) = \varphi(g(r)) \tag{6}$$

we will generalise (6) as follows:

$$g^*(r^*) = g^*(\psi(r)) = \varphi(g(r)) \tag{7}$$

so that also the source rankings are transformed. This very general model (7) will be studied in the next section and its equivalent size-frequency function $f^*$ will be calculated.

In Section III, the results obtained will be applied to Lotkaian systems and to transformations $\varphi$ and $\psi$ of power law type. Also in this case the equivalent size-frequency function $f^*$ will be calculated thereby extending the results in Egghe (2003, 2004) and Egghe and Rousseau (2006a).

Several applications of these results are described in Section IV. The applications go from general IPPs to countries or city size distributions, database distributions or (as initiated by V. Cothey) network distributions and their dynamics (evolutions).

# II. General evolutionary model for IPPs

Let us have a first system (IPP) given by $f:[a,\rho_m]\to\mathbb{R}^+,\ j\to f(j)$ as size-frequency function and by its equivalent (cf. (3), (4)) rank-frequency function $g:[0,T]\to\mathbb{R}^+,\ r\to g(r)$. Suppose this system is "changing" into a new system that we describe by asterisks: $f^*:\left[a^*,\rho_m^*\right]\to\mathbb{R}^+,\ j^*\to f^*\left(j^*\right)$ and $g^*:\left[0,T^*\right]\to\mathbb{R}^+,\ r^*\to g^*\left(r^*\right)$.

To allow for the largest possible freedom of evolution of the first IPP into the second we allow for a transformation of the source densities as well as of the item densities as follows: we define

$$g^*\left(r^*\right)=g^*\left(\psi(r)\right)=\varphi\left(g(r)\right) \tag{8}$$

where

$$\begin{aligned} \psi:[0,T]&\to\left[0,T^*\right] \\ r\to r^*&=\psi(r) \end{aligned} \tag{9}$$

is differentiable and where

$$
\begin{aligned}
\varphi : [a, \rho_m] &\to \left[a^*, \rho_m^*\right] \\
j &\to j^* = \varphi(j)
\end{aligned}
\tag{10}
$$

is differentiable.

Formula (8) describes the general rank-frequency transformation $g \to g^*$. The corresponding size-frequency transformation $f \to f^*$ is given by the next basic theorem.

**<u>Theorem II.1</u>**:

Formula (8) implies, for all $j \in [a, \rho_m]$ and $j^* \in \left[a^*, \rho_m^*\right]$:

$$
f^*\left(j^*\right) = f(j) \frac{\psi'\left(g^{-1}(j)\right)}{\varphi'(j)}
\tag{11}
$$

where $j^* = \varphi(j)$ as above (and assuming $\varphi' \neq 0$).

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

By the defining relation (3) we have

$$
r = g^{-1}(j) = \int_j^{\rho_m} f(k) dk
\tag{12}
$$

for all $j \in [a, \rho_m]$, $r \in [0, T]$ and

$$
r^* = g^{*-1}\left(j^*\right) = \int_{j^*}^{\rho_m^*} f^*\left(k^*\right) dk^*
\tag{13}
$$

for all $j^* \in \left[a^*, \rho_m^*\right]$, $r^* \in \left[0, T^*\right]$.

Hence, by (9), we have

$$\psi\left(\int_j^{\rho_m} f(k)dk\right) = \int_{j^*}^{\rho_m^*} f^*(k^*)dk^* \tag{14}$$

So, by (10)

$$\psi\left(\int_j^{\rho_m} f(k)dk\right) = \int_{\varphi(j)}^{\rho_m^*} f^*(k^*)dk^* \tag{15}$$

Differentiating both sides of (15) with respect to j yields:

$$\psi'\left(\int_j^{\rho_m} f(k)dk\right)(-f(j)) = -f^*(\varphi(j))\varphi'(j)$$

hence, by (10):

$$f^*(j^*) = f(j)\frac{\psi'\left(\int_j^{\rho_m} f(k)dk\right)}{\varphi'(j)}$$

which gives (11) by (12). W

**Corollary II.2**:

If $\psi = \mathrm{Id}$ (i.e. $\psi(r) = r$ for all $r \in [0,T]$) we have, for all j and $j^*$ as in Theorem II.1:

$$f^*(j^*) = \frac{f(j)}{\varphi'(j)} \tag{16}$$

Hence,

$$f^*(j^*) = \frac{f(\varphi^{-1}(j^*))}{\varphi'(\varphi^{-1}(j^*))} \tag{17}$$

**Proof**:

This is trivial since $\psi' = 1$ and by (10). W

This special case was already recovered in Egghe (2003, 2004) (for $a = a^* = 1$).

# III. Power law transformations in Lotkaian IPPs

Now the obtained results will be applied to Lotkaian IPPs where $\varphi$ and $\psi$ are transformations of power law type.

Lotkaian IPPs are IPPs where we have a decreasing power law for the size-frequency function f:

$$f(j) = \frac{C}{j^{\alpha}} \tag{18}$$

$C > 0$ and $\alpha > 1$ constants, $j \in [a, \rho_m]$.

## III.1 The case $\rho_m < \infty$

As proved in Egghe and Rousseau (1990) – see also Egghe (2005) – we now have for the rank-frequency function g (equivalent to f in (18)):

$$j = g(r) = \frac{E}{(1 + Fr)^{\beta}} \tag{19}$$

with

$$\beta = \frac{1}{\alpha - 1} \tag{20}$$

$$E = \rho_m \tag{21}$$

$$F = \frac{\alpha - 1}{C\rho_m^{1-\alpha}} \tag{22}$$

Note that the value of a is only implicitly involved in (19), being the lowest possible value for $g(r)$ (i.e. $a = g(T)$).

In this Lotkaian framework, we will also use (increasing) transformations $\varphi$ and $\psi$ of power type:

$$r^* = \psi(r) = Ar^b \qquad (r \in [0,T]) \tag{23}$$

$$j^* = \varphi(j) = Bj^c \qquad (j \in [a,\rho_m]) \tag{24}$$

with A, B, b, c > 0. We now have the transformation

$$g^*(r^*) = g^*(Ar^b) = B(g(r))^c \tag{25}$$

as follows from (8).

We will now evaluate the form of the transformed size-frequency function $f^*$. By (11) we have, since

$$\psi'(r) = Abr^{b-1} \tag{26}$$

$$\varphi'(j) = Bcj^{c-1} \tag{27}$$

that

$$f^*(j^*) = f(j)\frac{Ab\left(g^{-1}(j)\right)^{b-1}}{Bcj^{c-1}} \tag{28}$$

Since f is Lotkaian (18) we have (19) and (20) hence

$$r = g^{-1}(j) = \frac{\left(\frac{E}{j}\right)^{\alpha-1} - 1}{F} \tag{29}$$

Substituting (29) in (28) yields

$$f^*\left(j^*\right) = \frac{C}{j^{\alpha}} \; \frac{Ab\left[\left(\frac{E}{j}\right)^{\alpha-1} - 1\right]^{b-1}}{F^{b-1}Bcj^{c-1}}$$

$$f^*\left(j^*\right) = \frac{CAb}{F^{b-1}Bc} \; \frac{1}{j^{\alpha+c-1}} \left[\left(\frac{E}{j}\right)^{\alpha-1} - 1\right]^{b-1} \tag{30}$$

Now use (24) yielding

$$j = \left(\frac{j^*}{B}\right)^{\frac{1}{c}} \tag{31}$$

Formula (31) in (30) yields

$$f^*\left(j^*\right) = \frac{CAbB^{\frac{\alpha+c-1}{c}}}{F^{b-1}Bc} \; \frac{1}{j^{*\frac{\alpha+c-1}{c}}} \left[\frac{E^{\alpha-1}B^{\frac{\alpha-1}{c}}}{j^{*\frac{\alpha-1}{c}}} - 1\right]^{b-1}$$

Note that the expression between [ ], by (29) and (31) equals Fr.

So, for $r \in [0,T]$ large enough we have (since $Fr \approx Fr + 1$)

$$f^*\left(j^*\right) \approx \frac{CAbB^{\frac{\alpha-1}{c}}E^{(\alpha-1)(b-1)}B^{\frac{(\alpha-1)(b-1)}{c}}}{F^{b-1}c} \; \frac{1}{j^{*\delta}} \tag{32}$$

with

$$\delta = \frac{\alpha + c - 1 + (\alpha - 1)(b - 1)}{c} \tag{33}$$

and for $j^* \geq \varphi(a) = Ba^c$ (since $j \geq a$). Hence, denoting the intricate constant before $\frac{1}{j^{*\delta}}$ in (32) by G, we have, for large r (and for all r if $b = 1$)

$$f^*(j^*) \approx \frac{G}{j^{*\delta}} \tag{34}$$

with $\delta$ as in (33), i.e. Lotka's law with exponent $\delta$. Note that (34) is an equality if $b = 1$.So we have proved the following theorem.

**Theorem III.1**:
Let $\varphi$, $\psi$ and f be as in (18), (23) and (24). Then the transformed size-frequency function $f^*$ has the form

$$f^*(j^*) \approx \frac{G}{j^{*\delta}}$$

with $\delta$ as in (33), i.e. a Lotkaian size-frequency function where the exponent $\delta$ is the function (33) of the exponents b, c of the transformations $\psi$ and $\varphi$ respectively and of $\alpha$, the Lotka exponent of f and where $j^* \geq \varphi(a) = Ba^c$. If $b = 1$ then $\approx$ in (34) is an exact equality.

## III.2 The case $\rho_m = \infty$

Next we prove the same result for $\rho_m = \infty$ . We can now prove that (34) holds with an exact equality:

**Theorem III.2**:
Let $\varphi$, $\psi$ and f be as in (18), (23) and (24) with $\rho_m = \infty$ . Then the transformed size-frequency function $f^*$ has the form

$$f^*\left(j^*\right)=\frac{G}{j^{*\delta}} \tag{35}$$

with $\delta$ as in (33) and $j^* \geq \varphi(a)=Ba^c$.

**Proof**:

Instead of (19) we now have – see Egghe (2005), Exercise II.2.2.6 or Egghe and Rousseau (2006b) (Appendix) where a proof is provided:

$$g(r)=\frac{E}{r^{\beta}} \tag{36}$$

with $r \in \left]0,T\right]$, $E>0$ and $\beta$ as in (20). Since now

$$g^{-1}(j)=\left(\frac{E}{j}\right)^{\frac{1}{\beta}} \tag{37}$$

it follows from (11), (26), (27) and (37) that

$$f^*\left(j^*\right)=\frac{CAb\left(\frac{E}{j}\right)^{\frac{b-1}{\beta}}}{j^{\alpha}Bcj^{c-1}} \tag{38}$$

Now (31) yields, using (20)

$$f^*\left(j^*\right)=\frac{CAbE^{(b-1)(\alpha-1)}B^{\frac{\alpha-1}{c}b}}{cj^{*\delta}} \tag{39}$$

with $\delta$ as in (33), exactly and where $j^* \geq \varphi(a)=Ba^c$. □

We have the following trivial but important proposition.

**<u>Proposition III.2</u>**:

In the notation of above, we have

(i) $\delta < \alpha \Leftrightarrow b < c$

(ii) $\delta = \alpha \Leftrightarrow b = c$

(iii) $\delta > \alpha \Leftrightarrow b > c$

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

We only prove (i); the proof of (ii) and (iii) is similar. By (33):

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

iff

$$\alpha + c - 1 + (\alpha - 1)(b - 1) < \alpha c$$

iff

$$(\alpha - 1)(b - 1) < (\alpha - 1)(c - 1)$$

iff

$$b < c$$

since $\alpha > 1$. W

The interpretation of this corollary is important: Corollary III.2 gives necessary and sufficient conditions for the evolution of Lotkaian IPPs to result in higher or lower (or constant) Lotka exponents. In terms of sources and items this means, by (23), (24), that Lotka's exponent $\delta$ is decreasing under the transformation (i.e. $\delta < \alpha$) <u>if and only if</u> the "change" in the sources is smaller than the one in the items $(b < c)$. Analogous for the other assertions. Note also that if the transformations $\varphi$ and $\psi$ have the same exponents then $\delta = \alpha$, hence Lotka's exponent remains the same.

Note also that $\delta = \frac{c + (\alpha - 1)b}{c} = 1 + (\alpha - 1)\frac{b}{c}$. Hence $\delta$ only depends on $\alpha$ and the ratio of the exponents of the transformations $\varphi$ and $\psi$.

Summarising Section III, we have proved that power law transformations $\varphi$ and $\psi$ yield a Lotkaian IPP with exponent $\delta$ as in (33) if the original IPP is Lotkaian with exponent $\alpha$. As shown in (33), evidently, also the exponents of the power law transformations are involved. Proposition III.2 shows that $\delta$ and $\alpha$ relate as the exponents of the power law transformations in the sense that (in)equalities between $\delta$ and $\alpha$ are equivalent with similar (in)equalities between the exponents of the power law transformations.

In the next section we will discuss some (theoretical) applications.

# IV. Applications

IV.1 No sources are destroyed or created but one has that items can be destroyed (example: no nodes in a network are destroyed or created but one has the destruction of some in-links). Here $A = b = 1$ in (23), clearly. We can assume that the destruction of items follows a random sample in the items, hence sources with a large number of items have a higher probability for an item deletion, the probability being proportional to the source's size. This implies $c = 1$, $0 < B < 1$ in (24) (B being 1- sample probability (for destruction)). In this case we have $\delta = \alpha$ by (33) and (34) is an equality since $b = 1$. We hence refind Lotka's law with the same $\alpha$.

IV.2 No sources are deleted (destroyed) or created $(A = b = 1)$ but in a large source, items are deleted more than proportional to the source's size. Now we have $0 < c < 1$ and B must be choosen small enough to yield less items. Hence (33) yields $\delta > \alpha$. Indeed

$$\delta = \frac{\alpha + c - 1}{c} > \alpha \tag{40}$$

iff

$$\alpha + c - 1 > \alpha c$$

iff

$$(c-1)(\alpha-1) < 0$$

which is correct since $\alpha > 1$ and $c < 1$. Now we experience higher exponent values in Lotka's law (34) and again the result is exact since $b = 1$.

IV.3 No sources are destroyed or created $(A = b = 1)$ but items are destroyed preferably from low-item sources: $c > 1$ and B must be taken small enough (and certainly $< 1$) so that we have less sources. The same argument as in IV.2 now yields $\delta < \alpha$.

IV.4 The same result $(\delta < \alpha)$ was already found in Egghe (2003, 2004) in case of positive reinforcement (and again no sources are destroyed or created: $A = b = 1$): now $c > 1$ and B large enough to have more items. Again, as above, we have $\delta < \alpha$, as already found in Egghe (2003, 2004). In Egghe and Rousseau (2006a) the same model was used and in addition one supposed $B > 1$ yielding (see Theorem III.1), besides $\delta < \alpha$, that the minimal $j^*$-value $a^* = \varphi(a) = Ba^c$ is strictly larger than a, the minimal j-value in the original IPP. Repetition of this transformation leads to IPPs without low productive sources, hence explaining why these IPPs have smaller $(\delta < \alpha)$ Lotka exponents: see Egghe and Rousseau (2006a) for examples (cities/villages, countries and database sizes).

An extension of these results is obtained by considering other possible parameter values, but now leading to increased source and item totals.

IV.5 Let us have $0 < c \le 1$ (as in IV.1 and IV.2). Source deletion: let $b > 1$ and $A \in ]0,1[$ such that (see (23))

$$\psi(T) = AT^b = T^* < T$$

Then it is obvious from (33) and (40) that $\delta > \alpha$. The same is true for source creation ($b > 1$ and $A > 0$ such that $AT^b = T^* > T$).

IV.6 If $c > 1$ and if we have source destruction such that $b < 1$ (and A such that $\psi(T) = AT^b = T^* < T$) we have, by the argument in IV.3 and by (33) that $\delta < \alpha$. If $0 < c < 1$ we have no conclusion.

An example is given in Rosen and Resnick (1980) on the distribution of city sizes. Here one has the ambiguity of the definition of "city". One can use urban places, legal cities or urban agglomerations. In Rosen and Resnick (1980) one finds an increase of Zipf's exponent $\beta$ (there called the Pareto exponent) when going to the larger scale cities such as urban agglomerations. This boils down to a decrease of the Lotka exponent as indicated here $(\delta < \alpha)$. This is because of the inverse relation (20) (and similar for $\delta$).

In a way, this result generalizes IV.3 and IV.4 as well as Egghe (2003, 2004) and Egghe and Rousseau (2006a) (in the latter article the number of cities remains unchanged).

IV.7 Other examples of $\delta < \alpha$ or $\delta > \alpha$ can be constructed based on given parameter values A, B, b, c and giving further insight in the dynamics (evolution) of IPPs. We note, as in Egghe and Rousseau (2006a), that the given models of source/item creation present a non-stochastic form of general "Success-Breeds-Success" (SBS) principles.

# References


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, 2003, Dalian University of Technology Press, Dalian (China).

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

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

L. Egghe (2006). The theorem of Fellman and Jakobsson: a new proof and dual theory. Preprint.

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

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

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

J. Fellman (1976). The effect of transformations on Lorenz curves. Econometrica 44(4), 823-824.

U. Jakobsson (1976). On the measurement of the degree of progression. Journal of Public Economics 5, 161-168.

K. Rosen and M. Resnick (1980). The size distribution of cities: an examination of the Pareto law and primacy. Journal of Urban Economics 8, 165-186.