Solution of a problem of Buckland on the influence of obsolescence on scattering

Peer-reviewed author version

EGGHE, Leo (2004) Solution of a problem of Buckland on the influence of obsolescence on scattering. In: Scientometrics, 59(2). p. 225-232.

DOI: 10.1023/B:SCIE.0000018530.54281.68
Handle: http://hdl.handle.net/1942/739

Solution of a problem of Buckland on the influence of obsolescence on scattering

by

L. Egghe

Limburgs Universitair Centrum (LUC), Universitaire Campus, B-3590 Diepenbeek, Belgium[1]

and

Universiteit Antwerpen (UA), Universiteitsplein 1, B-2610 Wilrijk, Belgium

ABSTRACT

In an old paper [M.K. Buckland. Are obsolescence and scattering related? Journal of Documentation 28(3), 242-246, 1972] Buckland poses the question if certain types of obsolescence of scientific literature (in terms of age of citations) implies certain types of journal scattering (in terms of cited journals).

This problem is reformulated in terms of one- and two-dimensional obsolescence and linked with one- and two-dimensional growth, the latter being studied by Naranan. Naranan shows that two-dimensional exponential growth (i.e. of the journals and of the articles in journals) implies Lotka's law, a law belonging to two-dimensional informetrics and describing scattering of literature in a concise way.

In this way we obtain that exponential aging of journal citations and of article citations imply Lotka's law and a relation is given between the exponent α in Lotka's law and the aging rates of the two obsolescence processes studied.

[1] Permanent address

Key words and phrases: obsolescence, aging, growth, scattering, law of Lotka

I. Introduction

Obsolescence of literature is defined as the decreasing use of it in time, where "use" – in most cases – is expressed in terms of citations. Of course, citations refer to articles but can also be considered as citations to the journals in which these cited articles are published. It is then clear that use of less articles can lead to use of less journals in which these articles appear but this is not always the case. Indeed, take a typical citation-bibliography consisting of cited articles and journals. The informetrics of such a bibliography is – as always – very skew in the sense that few journals contain many citations, while many journals contain few (3, 2 or even 1) citations. Deleting a cited article, at random, from the bibliography (to indicate the diminishing use), probably, leads to deleting this article from a journal with many cited articles, since such journals have the highest probability to be encountered, and hence this process, in this case, does not lead to a decrease in the number of journals. Of course, it is always possible that the deleted article belongs to a journal that had only one cited article and in this case, this journal is also deleted in this process.

This obsolescence process, as described here, is completely equivalent with the growth process to be described now: let us have a situation where we have $t \in \mathbb{N}$ articles (here one can consider cited as well as published articles), scattered over a certain number of journals, where journals contain diverse quantities of articles. Note that we can interpret t (the total number of articles), equivalently, as the time t, to be used in the growth process. Let us add one more article (the $(t+1)^{th}$, illustrating growth). Then there is a certain chance p(n,t) that this new article will be published by an already existing source that had, at t, $n \in \mathbb{N}, n \le t$ articles but there is also a certain chance p(0,t) that this new article will be published by a new source (i.e. a source that, at t, did not have an article). Such mechanisms are completely determined by the respective probabilities p(n,t), n = 0,1,2,3,… and one usually refers to "success-breeds-success" (SBS) to describe such mechanisms. The mechanism is, in fact, wider than the describing name, where SBS refers to "the higher n, i.e. the higher the number of articles in an already existing source at t, the higher its probability to produce the $(t+1)^{th}$ article". Indeed, the mechanism is general enough to comprise other allocation schemes such as even "failure breeds failure". For more on SBS we refer to Simon (1955), De Solla Price

(1976), Egghe (1995, 1996, 2003a), Egghe and Rousseau (1990, 1995, 1996) and references therein.

The equivalence between the obsolescence process and growth process as described here is seen by applying the transformation $t' = -t$. Indeed, consider the obsolescence process where there is a decline of used articles when time increases (in the direction of the past). The transformation $t' = -t$ reverses time and makes us look at the future (with respect to a certain point in the past). Looking at the obsolescence process in this way we obtain a growth process

- of used articles in existing journals
- of journals itself, the probability of a new journal being the one for a deletion of a journal in the obsolescence process, as described above.

This equivalence will be formalized in the sequel.

The above description can be extended to the case of sources (generalizing journals) and items (generalizing (used) articles). The SBS-mechanism then tells us how a growth of the number of items is reflected in a growth of the number of sources. Without defining the exact mechanism (i.e. the probabilities $p(n,t)$, $n = 0,1,2,\ldots$, $t \in \infty$), one is not able to deduce the growth of the number of sources from the growth of the number of items and the same for the obsolescence of items and sources, because of the indicated equivalence. Since thus the source distribution does not follow from the item distribution we hence can talk about two-dimensional growth and obsolescence models, i.e. where growth (or obsolescence) is described by giving the source <u>and</u> item growth (or obsolescence) distribution.

Scattering can be defined as the distribution of items over sources, e.g. describing how many sources have how many items, i.e. describing the size-frequency distribution (such as e.g. the law of Lotka). Scattering is hence part of two-dimensional informetrics which is not only describing the number of items and the number of sources (this would be two times a one dimensional informetrics problem) but relates the sources with the number of items they <u>contain</u> (or produce). For more on this see e.g. Egghe and Rousseau (1990) or Egghe (2003a).

This brings us to the problem formulated by Buckland (1972): what is the relation between obsolescence and scattering? In Buckland (1972), this problem is discussed and reformulated as (using our definitions and terminology above): what is the relation between the one-dimensional obsolescence distribution of items and the one-dimensional obsolescence distribution of sources (the latter considered as scattering in Buckland (1972))? In view of the above considerations it is not possible to derive such a relation since it depends on the SBS-mechanism adopted. Remarkably, however, the graphs in Buckland (1972) reveal that Buckland, in his effort to formulate the problem in a concise way, is considering the obsolescence distributions of items (used articles, e.g. citations) and of sources (journals) in an independent way, occupying two axes of a plane, constituting two coordinates on which other variables might be dependent (see Fig. 1 in Buckland (1972)). Hence Buckland himself considers two-dimensional obsolescence.

Based on the above and inspired by the graphs in Buckland (1972) we can now reformulate the problem of Buckland as follows:

Problem of Buckland: Given two-dimensional obsolescence (i.e. the aging distributions of the items as well as of the sources), can one determine scattering in the sense above: two-dimensional informetrics e.g. expressed by the size-frequency distribution?

The answer to the above question is yes, as we will explain in a mathematically concise way in the next section. We will now indicate below how we will proceed. First of all both obsolescence distributions are transformed (as indicated above) to growth distributions. In case we assume the obsolescence distributions to be decreasing exponential distributions (the simplest model and conforming with the graphs in Fig. 1 in Buckland (1972)), we then obtain increasing exponential growth distributions (for sources as well as for items). Then we recall an important result of Naranan (1970) (reproved and linked with the theory of self-similar fractals in Egghe (2003b), see also Egghe (2003a)), stating that if sources grow exponentially and if items in sources grow exponentially (with the same growth rate in every source) then the size-frequency distribution, describing the number of sources with n Î ¥ items is Lotkaian, i.e. a function of the form

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

where C, $\alpha > 0$. The theorem also clarifies the relation between Lotka's exponent α and the two growth rates (exact relations will be given). Because of the transformation of the obsolescence problem into a growth problem (in an equivalent way) we will, hence, obtain the new result that exponential obsolescence distributions for sources and items imply a Lotkaian size-frequency distribution and where we have an exact relation between Lotka's exponent α and the given obsolescence rates. This can be considered as the answer to the problem that was (implicitely) formulated by Buckland in 1972.

This paper closes with the link of this theory to the theory of self-similar fractals in the same way as was done in Egghe (2003b) (using two-dimensional growth processes).

II. Solution of the problem of Buckland

In the previous section we used discrete time $t \in \mathbb{N}$ in order to give an intuitively more appealing description of what we want to study in this paper. For the mathematically correct description of the relation between obsolescence and scattering we need continuous time $t \in \mathbb{R}^+$.

We will continue using the terminology of items and sources but it is handy to interpret the framework in terms of used (e.g. cited) articles (as items) and of the journals in which these used articles are published. Two-dimensional obsolescence implies that we give the source- and item-obsolescence distribution. In this paper we will work with the simplest models, being basic simplifications of other models (see also Egghe and Rao (1992)), namely decreasing exponential distributions (or rather functions giving actual quantities).

So we base ourselves in the present (representing time t = 0) and we look into the past (denoted by time t > 0). The decrease of used articles (as items) and of corresponding journals (as sources) is expressed by two decreasing exponential functions described as follows:

(i) The number of sources decreases exponentially in time $t \in [0, t_0]$ in the past. Heuristic (discrete) interpretation: t_0 is the time (in the past) when the last source “disappears”:

$$N(t) = c_1 b_1^{t - t_0} \tag{2}$$

where $c_1 > 0$ and $0 < b_1 < 1$ (to make sure that N is decreasing) are parameters.

(ii) The number of items in each source decreases exponentially in time $t \in [0, t_1]$ in the past. Heuristic (discrete) interpretation: t_1 is the time in the past when the source disappears (according to (i): i.e. when the last item in the source disappears):

$$q(t) = c_2 b_2^{t - t_1} \tag{3}$$

where $c_2 > 0$ and $0 < b_2 < 1$ are parameters, assumed to be the same for every source (a simplifying assumption) (t_1 is variable over the sources, of course).

Note that

$$\begin{aligned} N(t) &= c_1 b_1^{t - t_0} \\ &= c_1 \left(\frac{1}{b_1}\right)^{t_0 - t} \\ &= c_1 \left(\frac{1}{b_1}\right)^{t'} \\ &= c_1 a_1^{t'} \\ &=: M(t') \end{aligned} \tag{4}$$

where $c_1 > 0,\ \ a_1 = \dfrac{1}{b_1} > 1$ are parameters and where $t' =: t_0 - t \in [0, t_0]$ increases into the future when t decreases into the past. Similarly

$$
\begin{aligned}
q(t) &= c_2 b_2^{t-t_1} \\
&= c_2 \left(\frac{1}{b_2}\right)^{t_1 - t} \\
&= c_2 \left(\frac{1}{b_2}\right)^{\tau} \\
&= c_2 a_2^{\tau} \\
&=: m(\tau) \qquad (5)
\end{aligned}
$$

where $c_2 > 0,\ \ a_2 = \dfrac{1}{b_2} > 1$ are parameters and where $\tau =: t_1 - t \in [0, t_1]$ increases into the future when t decreases into the past; in fact, τ is the age of the source.

Making now the present arbitrary, we hence arrived at the following situation:

(a)

$$M(t') = c_1 a_1^{t'} \qquad (6)$$

$c_1 > 0,\ \ a_1 > 1$ is the formula for the number of sources at time $t' > 0$.

(b)

$$m(\tau) = c_2 a_2^{\tau} \qquad (7)$$

, $c_2 > 0,\ \ a_2 > 1$ is the formula for the number of items in a source of age $\tau > 0$ and c_2 and a_2 are (assumed to be) the same for every source.

We now invoke a theorem of Naranan (1970), reproved in Egghe (2003b), see also Egghe (2003a)) on the relation of two-dimensional growth and scattering:

Theorem II.1 (Naranan):
Under the assumptions (a) and (b) above, we have that

$$f(j) = \frac{C}{j^{\alpha}} \tag{8}$$

for $j > 0$, where f(j) denotes the size-frequency distribution, i.e. f(j) is the density of sources with item-density j, i.e. Lotka's law. Furthermore, the Lotka parameter α satisfies

$$\alpha = 1 + \frac{\ln a_1}{\ln a_2} \tag{9}$$

The above arguments now yield the following new result, establishing a relation between two-dimensional obsolescence and scattering.

Theorem II.2:
Let us have a process in which one has

(i) The number of sources decreases exponentially in time $t \in [0, t_0]$ in the past:

$$N(t) = c_1 b_1^{t - t_0}$$

where $c_1 > 0$ and $0 < b_1 < 1$ are parameters.

(ii) The number of items in a source decreases exponentially in time $t \in [0, t_1]$ in the past:

$$q(t) = c_2 b_2^{t - t_1}$$

where $c_2 > 0$ and $0 < b_2 < 1$ are parameters.

Then, if f(j) denotes the size-frequency distribution, describing the density of sources with item-density j, we have Lotka's law

$$f(j) = \frac{C}{j^{\alpha}} \tag{10}$$

$j > 0$ and where Lotka's α is given by

$$\alpha = 1 + \frac{\ln b_1}{\ln b_2} \tag{11}$$

Proof:
This follows readily from Theorem II.1, using formulae (4) and (5). Formula (11) follows from formula (9) by using that $a_1 = \frac{1}{b_1}$ and $a_2 = \frac{1}{b_2}$. ~

Corollary II.3:
If the obsolescence rates b_1 and b_2 are equal, then Lotka's $\alpha = 2$.

This closes the logical reformulation of the problem of Buckland as well as its solution.

III. Two-dimensional growth or obsolescence and the link with self-similar fractals

In Egghe (2003b), based on Theorem II.1 of Naranan, one could show that systems as described in Theorem II.1 can be considered as self-similar fractals with fractal dimension D being equal to

$$D = \frac{\ln a_1}{\ln a_2} \tag{12}$$

hence, using formula (9)

$$D = \alpha - 1. \tag{13}$$

Reformulated in terms of the obsolescence aging rates b_1 and b_2 we hence have, using (11) that

$$D = \alpha - 1 = \frac{\ln b_1}{\ln b_2} \tag{14}$$

Intuitively, a self-similar fractal (in a k-dimensional space, k = 1,2,3,…) is a subset of $\mathbb{R}^k$ such that, when we reduce the set by a factor $0 < \rho < 1$, we need $N > 1$ times this reduction in order to cover the original set and then its fractal dimension is given by

$$N = \frac{\ln N}{-\ln \rho} \tag{15}$$

The fractal dimension is a measure of "how spread out" the fractal is – see Egghe (2003a) for a more detailed description – and hence is a measure of its complexity. Formula (12) (and hence also formula (14)) is obtained by applying the reduction $b_2^t = \left(\frac{1}{a_2}\right)^t$ (the number of items in a source expresses the scale at which we look at the process) and then expressing that we need $a_1^t = \left(\frac{1}{b_1}\right)^t$ sources (times the number c_1 = the number of sources at t = 0). As in Egghe (2003b) we have the following corollary:

<u>Corollary III.1</u>:

The obsolescence rates b_1 and b_2 are the same then we have a self-similar fractal with fractal dimension equal to 1.

This follows readily from formula (14).

References

M.K. Buckland (1972). Are obsolescence and scattering related? Journal of Documentation 28(3), 242-246, 1972.

L. Egghe (1995). Extension of the general “success-breeds-success” principle to the case that items can have multiple sources. Proceedings of the fifth biennial Conference of the international Society for Scientometrics and Informetrics, River Forest (USA) (M. Koenig and A. Bookstein, eds.), 147-156, 1995. Learned Information, Medford (USA).

L. Egghe (1996). Source-item production laws for the case that items have multiple sources with fractional counting of credits. Journal of the American Society for Information Science 47(10), 730-748, 1996.

L. Egghe (2003a). Lotkaian Informetrics. Book in preparation.

L. Egghe (2003b). The power of power laws and an interpretation of Lotkaian informetric systems as self-similar fractals. Preprint 2003.

L. Egghe and I.K. Ravichandra Rao (1992). Citation age data and the obsolescence function: fits and explanations. Information Processing and Management 28(2), 201-217, 1992.

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

L. Egghe and R. Rousseau (1995). Generalized success-breeds-success principle leading to time dependent informetric distributions. Journal of the American Society for Information Science 46(6), 426-445, 1995.

L. Egghe and R. Rousseau (1996). Stochastic processes determined by a general success-breeds-success principle. Mathematical and Computer Modelling 23(4), 93-104, 1996.

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

H.A. Simon (1955). On a class of skew distribution functions. Biometrika 42, 425-440, 1955.

D. De Solla Price (1976). A general theory of bibliometric and other cumulative advantage processes. Journal of the American Society for Information Science 27, 292-306, 1976.