Aging, obsolescence, impact, growth and utilization: definitions and relations

Peer-reviewed author version

EGGHE, Leo & ROUSSEAU, Ronald (2000) Aging, obsolescence, impact, growth and utilization: definitions and relations. In: Journal of the American Society for Information Science, 51(11). p. 1004-1017.

DOI: 10.1002/1097-4571(2000)9999:9999<::AID-ASI1003>3.0.CO;2-8
Handle: http://hdl.handle.net/1942/786

Aging, obsolescence, impact, growth and utilization: definitions and relations

Leo Egghe*[*] and Ronald Rousseau[**]

[*] LUC, Universitaire Campus, B-3590 Diepenbeek, Belgium

and UIA, IBW, Universiteitsplein 1, B-2610, Wilrijk, Belgium

leo.egghe@luc.ac.be

[**] KHBO, Department of Industrial Sciences and Technology, Zeedijk 101, B-8400 Ostend, Belgium

and UIA, IBW, Universiteitsplein 1, B-2610, Wilrijk, Belgium

ronald.rousseau@kh.khbo.be

Abstract

The notions aging, obsolescence, impact, growth, utilization and their relations are studied. It is shown how to correct an observed citation distribution for growth, once the growth distribution is known. The relation of this correction procedure with the calculation of impact measures is explained. More interestingly, we have shown how the influence of growth on aging can be studied over a complete period as a whole. Here the difference between the so-called average and global aging distributions is the main factor. Our main result is that growth can influence aging but that it does not cause aging. A short overview of some classical articles on this topic is given. Results of these earlier works are placed in the framework set up in this article.

Keywords and phrases: age, average and global aging distributions, growth, reversed growth, utilization, obsolescence, impact, growth rates, temporal selectivity

1. Introduction

The terms 'growth' and 'to grow'

Aging and growth are two important topics in information science. Knowledge of these concepts leads to the study of the evolution of scientific fields and allows comparisons between fields. Yet it is sometimes not clear what is meant by the terms used in such studies. A point in case is the meaning of the notions 'growth' and 'to grow'.

Sometimes the terms 'growth' and 'to grow' are meant in the sense of the number of entities that come into existence, e.g. articles published, during a fixed period (say one year). Thus 'yearly growth' means the same as 'yearly production' of entities. We will use these terms mainly in this sense. On other occasions the terms 'growth' and 'to grow' refer to the total (cumulative) production over many years, as in the phrase 'this year the database has grown to a total of 1 million items'. Finally, the term 'growth' is sometimes used in the sense of an expanded activity, i.e. the (positive) difference of the production between consecutive moments in time, as in the phrase, 'the growth of the GNP' (gross national product). We have the impression that when information scientists studied the 'influence of growth on obsolescence' (to be defined shortly) they used the term 'growth' mostly in the latter sense.

Conglomerates, sources and items

As in many earlier publications (Egghe & Rousseau, 1990) we use the general terms sources and items. Sources are themselves units in a larger set and come into existence through an information production process. This larger set may be a bibliography on a subject or a publication list of a scientist of group of scientists. As a generic name for this larger set we used the term 'generalized bibliography' in earlier publications. Perhaps a new, and more neutral term, such as 'conglomerate' would be appropriate.

We stress the fact that in studying aging effects time is clearly an important and complicating factor. In general we consider a conglomerate as a dynamic entity (but static ones are not excluded). As always sources (can) produce items. To clarify what we mean by these concepts we first present three examples.

1°) Compiling a bibliography is one way to represent a scientific discipline. Such a bibliography comes only gradually into existence (it grows). It is an example of a conglomerate. Each source of the conglomerate is a published article and each time one of these articles is cited, it is said that this article has produced an item. So, in this example articles in the bibliography are sources. The total of all articles, i.e. the conglomerate, grows and a citation to one of these articles is an item. Instead of a discipline, we may, similarly, study the lifetime scientific achievements of a scientist or a group of scientists.

2°) A totally different situation occurs in the study of Internet pages with, e.g., *.be* in its URL (Belgian web pages). Again this conglomerate changes in the course of time, but it does not necessarily increase (in principle it may even decrease). Each such page is a source, and each time another web page makes a link to such a *.be*-page an item is created (Rousseau, 1997). Note that the link is the item, not the other web page. The disappearance of a web page makes all produced items without 'content'. They are considered to disappear together with the source.

3°) It is also possible to study a fixed set of articles, e.g. one issue of a journal. There is no growth in time, but all other aspects stay the same. The collection of articles published in that issue forms the (static) conglomerate, the articles themselves are sources and each time one of these articles is cited, an item is created (cf. example 1°).

Aging - obsolescence

Although the term 'obsolescence' has negative connotations, it is a commonly used notion. We would describe it as the (possible) decline of usefulness over time. In this article we tend to avoid this negative term and use 'aging' instead. In the case of published articles aging is often measured by counting citations. Obsolescence can also be described as temporal selectivity in the use of articles. For us the notions 'aging curve and aging distribution' are synonymous with the notions 'citation curve and citation distribution'. In the first months after publication aging of an article really means an increase in citations. In that period we have a 'burgundy effect' (becoming better with age) (Degroote et al., 1991). In relation with the second example we remark that removal of a page is the ultimate form of obsolescence! For a review on concrete aging distributions of citations we refer the reader to (Egghe & Rao, 1992a; Matricciani, 1991; Egghe & Rousseau, 1990). Use can also be considered in a relative way, namely use with respect to supply. A typical example is the number of citations with respect to the number of publications, a notion that is usually referred to as 'impact'. Thus impact can be considered as a form of relative aging. Moreover, we will show that 'correcting aging for growth' is closely related to studying impact.

Time as a discrete variable

The term 'aging' leads us again to the notion of time. Time is a continuous variable. Yet one is often interested in variations between numbers measured on a fixed time scale, usually one year. Moreover, data is often collected in a discrete way. Hence, time will be treated as a discrete variable. We would like to point out though that most of our results are equally valid when considering time as a continuous variable.

Diachronous and synchronous studies

Aging can be studied in a synchronous and in a diachronous way (Nakamoto, 1988). Diachronous studies of aging consider a source (or a set of sources) and examine the number of items produced by this source later on. One studies e.g. a publication and the number of citations it receives over the years. If the original source (publication) comes into existence at this moment, one has to wait a long time before a diachronous study becomes meaningful. In principle a diachronous study never ends. On the other hand, synchronous studies look at the past. One examines e.g. the reference list of a recently published article (or group of articles) and determines the ages of the references in that (or these) article(s). Diachronous studies (of articles) require citation indices, while synchronous studies do not. Hence synchronous studies are usually cheaper and easier to perform. According to Stinson (1981), see also (Stinson & Lancaster, 1987), synchronous and diachronous studies of the aging of scientific articles lead to the same conclusions, hence implying a preference for synchronous ones.

The three groups of examples of conglomerates, sources and items mentioned above are all described in a diachronous way. They may also be studied in a synchronous way (see Table 1). We further note that synchronous as well as diachronous studies may be done 'in retrospect', by which we mean that such a study is being done (now) with a moment in the past as starting or reference date. In such a way one may perform in the year 2001 a diachronous as well a synchronous study of articles published, e.g. in the year 1990. The main difference between the diachronous and the synchronous approach is that in the synchronous approach sources 'produce' items that existed already, i.e. the sources are active in establishing the source-item relation. A typical example is the case of an article (the source) citing another (necessarily older) one (the item). In the diachronous approach the items establish the source-item relation. A typical example here is that of an article (the source) being cited by another (necessarily younger) one (the item).

Although in practical cases diachronous and synchronous studies may need different techniques, this distinction plays no role on the (theoretical) level we adopt in the present study. Our notation is so general as to incorporate both the diachronous and the synchronous case.

More examples

Table 1 gives a number of practical examples of conglomerates, sources and items.

[insert Table 1 about here]

Growth functions and rates of growth

As stated above, the time evolution of the conglomerate (set of sources) under consideration will be described through a growth function, denoted as g(t). Here the variable t denotes time and is considered as increasing from the past to the present. In most cases this growth function itself is increasing, but we do not exclude a constant or decreasing growth function. We will, however, also have to 'go back in time'. In that case we will use the term 'reversed growth distribution', denoted as j(t). The variable t increases from the present to the past. For a review of concrete growth functions such as the exponential function or S-shaped ones, we refer the reader to (Tague et al., 1981; Egghe & Rao, 1992b).

Rates of growth or aging will be determined by the value at time t+1 divided by the value at time t. Hence the (discrete) growth rate will be given by g(t+1)/g(t) and the (discrete) aging rate by c(t+1)/c(t) (Egghe & Rao, 1992a,b). The main purpose of this article is to study the disturbing role of growth on aging and utilization (see further for a description of this term). Note that sometimes the term 'growth rate' refers to the logarithm of what we have described as the 'growth rate' (Pierou et al. 1999), or to the expression (g(t+1)-g(t))/g(t) (see e.g. (Narayana, 1998)).

Table 1

dynamic (or static) conglomerate	**sources**	**items**
Synchronous examples		
scientific discipline, represented by a bibliography (dynamic case)	published documents	references in these documents
one issue of a journal (static case)	articles in this issue	references in these articles
fiction literature	books	words used in these books
Internet presence of a country, as represented by URLs	pages with this country's code in the URL	links in these pages (to other pages)
scientific output of a research group	published documents	references in these documents
Diachronous examples		
scientific discipline, represented by a bibliography	documents published from now on	citations of these documents
scientific output of a research group	published documents	citations of these documents
one issue of a journal	articles in this issue	citations of these articles
public or scientific library	acquired books (from now on)	loans of these new books
Internet presence of a country, as represented by URLs	pages with this country's code in the URL	links to these pages
database	records	searches that retrieve these records

Applications

Obsolescence and aging have often been studied as an aspect of a particular literature or journal (Clark, 1976; Diodato & Smith, 1993; Rowlands, 1999). They have also been studied in the framework of finding the best possible policy for relegating rarely used library items to secondary storage (Brookes, 1970). Scales' (1976) and Tsay's (1998) articles are among the few ones trying to link library use with global use as measured through citations. Griffith et al. (1979) have shown that scientific journals used as archives age slowly, while journals supporting the research front age quickly. Our article concentrates on the methodological issues and the relations between the different notions involved in this kind of study.

Outline of the article

The next section presents the influence of growth on aging, how aging rates can be corrected for growth and the relation with impact measures. Here our attention goes essentially to a single one-year period. In the third and fourth sections we give our attention to the case that a whole period of length T is studied at once. A number of general (mathematical) results will be proven in section 5. The sixth section provides practical calculations and examples of the concepts studied in the previous sections. Section 7 gives a short overview of older articles on this topic. Results are placed in the framework set up in this article. Finally, we come to the conclusions.

2. The reversed growth distribution and its influence on aging

2.1 The basic distributions

To fix the ideas our descriptions will be cast in a synchronous framework of a scientific discipline, studied through a bibliography of publications and references made in these publications. Note though that our results apply equally well to other source-item relations, as well as to the diachronous case.

For every $t \in \mathbb{N}$ we denote by c_t the observed discrete distribution function (in the statistical sense) of the publication data of the references in the documents in the bibliography that were published t years ago. The function $c_t(n)$ is called the observed instantaneous aging distribution. Note that we use a symbol such as n to denote the variable of the function c_t, with $n \in \mathbb{N}$. Hence, n increases in the direction of the past, starting at t (where n = 0). See Fig.1.

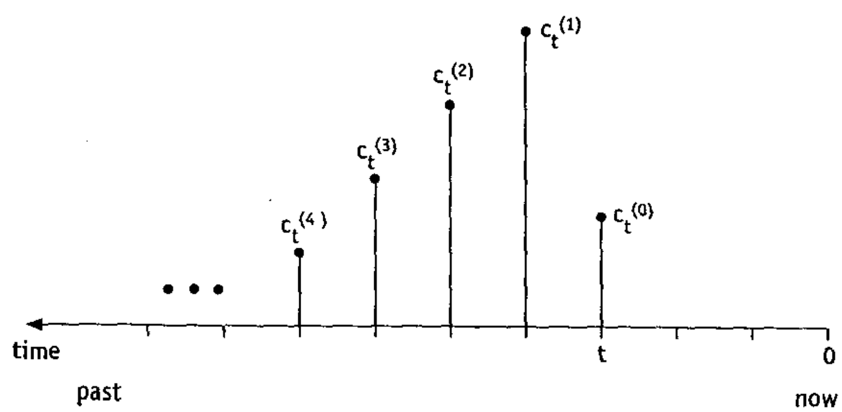

Fig. 1 Observed instantaneous aging distribution on time axis

As distribution functions the c_ts satisfy the relation: $\forall$ t $\in \mathbb{N}$

$$\sum_{n=0}^{\infty} c_t(n) = 1 \qquad (1)$$

We further introduce a distribution function that describes how the bibliography grows. For technical reasons it is at this moment more convenient to describe such a function with time t increasing from now (point 0) in the direction of the past. This function is denoted as j(t) and is referred to as the reversed growth distribution function. Note that the (usual) growth function, g(t) is essentially the same function as the reversed growth function but now time is moving in the natural direction, i.e. from the past to the future (see further Section 3 for a more precise definition of a growth function). Of course, we also have:

$$\sum_{t=0}^{\infty} j(t) = 1 \qquad (2)$$

The reversed growth rate function $\chi(t)$ is defined as:

$$\chi(t) = \frac{j(t+1)}{j(t)} \qquad (3)$$

The change in production between moments that are separated by more than one year is denoted and defined as follows, for t= 0,... and n= 1, 2,...:

$$\chi(t,n) = \frac{j(t+n)}{j(t)} \qquad (4)$$

Note that

$$\chi(t,n) = \chi(t) \cdot \chi(t+1) \cdots \chi(t+n-1) \ \textit{and that} \ \chi(t,1) = \chi(t). \qquad (5)$$

We put $\chi(t,0) = 1$. If the growth is constant then $\chi(t,n) = 1$ for every t and n. If the reversed growth function declines exponentially (corresponding with an exponential growth): $j(t) = e^{-at}$, then $\chi(t)$ is constant (= e^{-a}). Note that a = 0 yields the case of constant growth.

Citation distributions and (reversed) growth curves may be determined for real bibliographies. An obvious question: what is the simplest way to take the growth of the literature into account when studying citations? To answer this question in a specific year (say now, i.e. the year 0) one must replace

$$c_0(n) \text{ by } k_0(n) = K_0 \frac{c_0(n)}{\chi(0,n)} \qquad (6)$$

where $k_0(n)$ denotes the corrected instantaneous citation distribution, and K_0 is a normalization factor introduced to make k_0 into a distribution function. Assume, e.g., that the literature has increased by 10% over the past 5 years. Then $\chi(0,5)$ = 100/110 ($\approx$ 0.91) and the observed citations are multiplied by 110/100 = 1.10. It is clear that if growth is constant, the corrected citation curve becomes equal to the observed one: no correction is necessary. In general, we define

$$k_t(n) = K_t \frac{c_t(n)}{\chi(t,n)} \qquad (7)$$

with K_t as a normalization factor.

For every $t \in \mathbb{N}$, t fixed, the instantaneous observed aging rate function is defined as:

$$\beta_t(n) = \frac{c_t(n+1)}{c_t(n)} \qquad (8)$$

with $n \in \mathbb{N}$. Similarly, the instantaneous corrected aging rate is:

$$\rho_t(n) = \frac{k_t(n+1)}{k_t(n)} \qquad (9)$$

Note that:

$$\rho_t(n) = \frac{\beta_t(n)}{\chi(n+t)} \qquad (10)$$

The meaning of the reversed growth rate function is clear: one considers, for every t in the period under study, the number of publications that originated t+1 time units (years) ago, divided by the number of publications that originated t years ago. The instantaneous aging rate functions $\beta_t(n)$ and $\rho_t(n)$ can be interpreted as follows. At each time t we consider all publications in the bibliography that were published at that moment (or that year). The age distribution of the references of these publications (corrected for growth or not) is given by the functions k_t and c_t. We take the ratio of the number of references that had been published n+1 year earlier (hence t + n +1 year from now) and the number of references that had been published n years earlier (hence t + n year from now). If the growth rate is constant (exponential growth) the corrected and non-corrected aging rates differ by this fixed constant. If, however, the growth function is a constant function then $\chi = 1$ and hence $\rho = \beta$.

Based on Equation (10) we can make the following comments. If the reversed growth rate is always smaller than one: $\chi(t) < 1$, then $\beta_t(n) < \rho_t(n)$, for every t and n. In particular, if the corrected aging rate is smaller than one this implies that the observed aging rate is even smaller. So growth of the literature increases the obsolescence effect. If the observed aging rate is larger than one (period in which the burgundy effect prevails) the corrected one is even larger. So, if the observed citation curve is unimodal (say roughly lognormally distributed) the above observations mean that the mode of the corrected curve is shifted towards larger t-values. So, *a decreasing reversed growth (i.e. increasing growth) intensifies the immediacy effect.* This is shown in Fig.2. For clarity's sake we have used continuous distribution functions in this illustration.

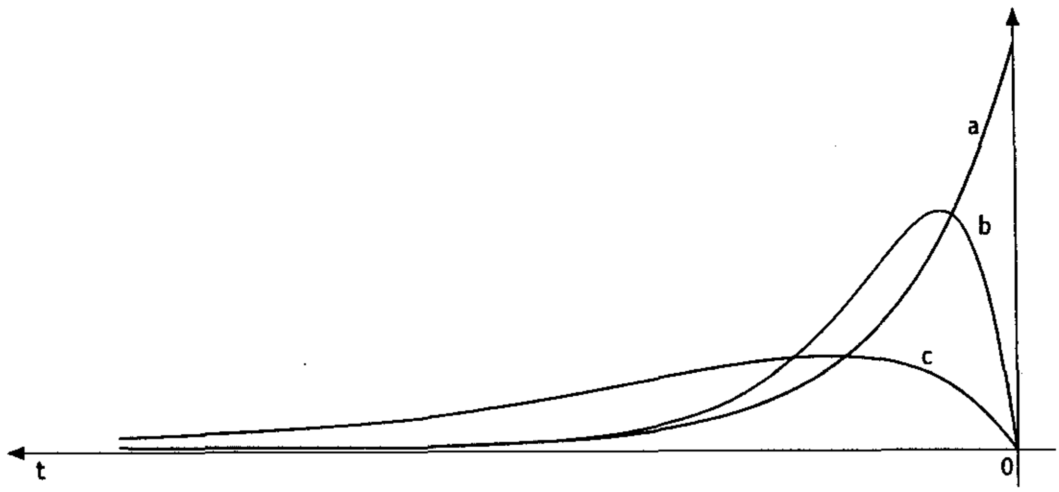

Fig. 2 Increasing growth intensifies the immediacy effect
a. reversed growth distribution
b. observed aging distribution
c. corrected aging distribution

One could also consider instantaneous impact factors, denoted as $i_t(n)$, and defined as:

$$i_t(n) = \frac{c_t(n)}{j(n+t)} \qquad (11)$$

If the observed citation curve exactly followed the reversed growth curve, the instantaneous impact factors would be constant. One could say that in this situation there is no relative obsolescence. Table 2 gives an example of a situation where this is the case.

Table 2 An example where the instantaneous impact curve is always constant.

N	0	1	2	3	4	5
j(n)	320	160	80	40	20	10
$c_t(n)$						
t = 0	640	320	160	80	40	20
t = 1	320	160	80	40	20	...
t = 2	160	80	40	20	...	
t = 3	80	40	20	...		
t = 4	...					
$i_t(n)$						
t = 0	2	2	2	2	2	2
t = 1	2	2	2	2	2	
t = 2	...					

The following proposition shows that correcting a citation distribution for the growth of the literature is exactly the same as considering impact factors.

Proposition 1

The corrected instantaneous citation distribution is the same as the instantaneous impact factor distribution.

Proof

We know that, for fixed t,

$$k_t(n) = K_t \frac{c_t(n)}{\chi(t,n)} \qquad (7)$$

On the other hand, multiplying (11) by an appropriate factor in order to obtain a distribution function leads to:

$$i_t(n) = I_t \frac{c_t(n)}{j(n+t)} \qquad (12)$$

where we have kept the same notation for impact. However, by Equation (4), we have:

$$k_t(n) = K_t \frac{c_t(n)\, j(t)}{j(t+n)} = I_t \frac{c_t(n)}{j(t+n)} = i_t(n) \quad (13)$$

with $K_t\, j(t) = I_t$. This proves Proposition 1.

Therefore we note that the instantaneous corrected aging rate is equal to the instantaneous impact rate.

Proposition 2

If χ is decreasing and if the instantaneous observed aging rate functions β_t are strictly increasing in t, then the instantaneous corrected aging rate functions ρ_t are strictly increasing in t.

Proof: this result follows from:

$$\beta_t(n) < \beta_{t+1}(n) \Leftrightarrow \rho_t(n)\cdot\chi(n+t) < \rho_{t+1}(n)\cdot\chi(n+1+t)$$
$$\Rightarrow \rho_t(n)\cdot\chi(n+t+1) < \rho_{t+1}(n)\cdot\chi(n+1+t) \Leftrightarrow \rho_t(n) < \rho_{t+1}(n)$$

Similarly we obtain:

If χ is increasing and if the instantaneous observed aging rate functions β_t are strictly decreasing in t, then the instantaneous corrected aging rate functions ρ_t are strictly decreasing in t.

The comments made about Equation (10) concerning the observed and the corrected aging curves and aging rates, may, by Proposition 1, also be made concerning the observed aging curve and rates and the impact.

With the functions c_t and j we can define (generalized) impact factors, denoted as IF. We can consider, e.g.:

$$IF_t(n) = \frac{\sum_{k=0}^{n} c_t(k)}{\sum_{k=0}^{n} j(k+t)} \qquad (14)$$

These impact factors were introduced in (Rousseau, 1988) and studied, in a continuous setting, by Egghe (Egghe, 1988). We note (and prove) one result as an example.

Proposition 3

For all $n_0 \in \mathbb{N}_0$: $IF_t(n_0) = IF_t(n_0 - 1)$ if and only if $i_t(n_0) = IF_t(n_0)$.

Proof. If $i_t(n_0) = IF_t(n_0)$, then

$$c_t(n_0)\big(j(t)+\cdots+j(n_0+t)\big)=j(n_0+t)\big(c_t(n_0)+\cdots+c_t(0)\big)$$

Consequently, we also have:

$$c_t(n_0)\big(j(t)+\cdots+j(n_0-1+t)\big)=j(n_0+t)\big(c_t(n_0-1)+\cdots+c_t(0)\big)$$

Hence: $i_t(n_0) = IF_t(n_0 - 1)$, and, thus $IF_t(n_0) = IF_t(n_0 - 1)$. The other implication may be shown in a similar way.

This ends our observations about the influence of reversed growth on aging if one is mainly interested in one particular year, i.e. c_t, with t fixed. Most of what we have written here is well known (see Section 7). In the other sections of this article we study what happens if one considers a longer time period, i.e. considering several t-values at once. To a great extent this is a new aspect of aging.

3. Definitions and properties of average age, aging and utilization

Suppose we consider a bibliography, i.e. a set of documents, published over a certain time period, denoted as [0,T]. Here, the symbol 0 (zero) represents the present and T refers to the moment of publication of the oldest document in the collection. Of course, there may exist documents older than T, but these are not considered as sources. They may, however, play the role of items, i.e. it is possible to cite articles older than T.

3.1 The growth function

We denote by g(t), t ∈ [0,T], the discrete growth distribution function in the investigated period [0,T]. For the growth function g, time increases from the oldest document in the collection under study (which came into existence at time 0) in the direction of the present (time T). See Fig.3

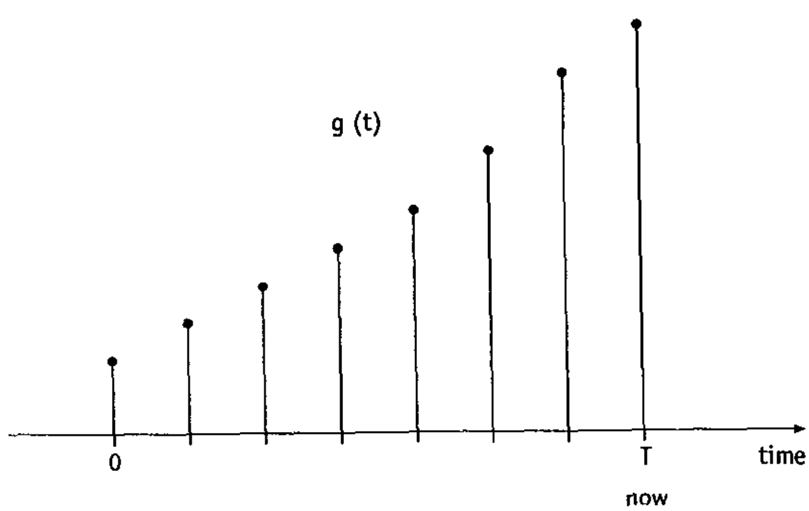

Fig.3 Growth function

We note that also g, being a distribution function, satisfies:

$$\sum_{t=0}^{T} g(t)=1 \qquad (15)$$

If j(t) = 0 for t > T, then, for all t ∈ [0,T]:

$$j(t) = g(T-t) \qquad (16)$$

For simplicity's sake we assume from now on that (16) is always satisfied.

The growth rate function of the bibliography is, for t ∈ [0,T-1], defined as:

$$\alpha(t) = g(t+1) / g(t) \qquad (17)$$

Then:

$$\alpha(t) = \frac{1}{\chi(T-t-1)},\ or\ \ \chi(t) = \frac{1}{\alpha(T-t-1)} \qquad (18)$$

The instantaneous aging rates studied in the previous section only present information about aging at a fixed time t. It is much more important to define aging functions that measure aging of the bibliography as a whole (hence considering the whole period [0,T]). To this end we introduce a number of functions. Note that, although we work with the observed aging distributions $c_t(n)$, we could (and perhaps better) use the corrected aging distributions $k_t(n)$ as well. At this point this is, however, only a matter of notation.

3.2 The average aging distribution function of a bibliography (in short: the aging distribution), denoted as c_A.

The function $c_A(n)$, n ∈ ℕ, is defined as the average of the instantaneous aging distributions c_t over t ∈ [0,T]:

$$c_A(n) = \frac{1}{T+1} \sum_{t=0}^{T} c_t(n) \qquad (19)$$

It is easy to check that the function c_A is itself a distribution function:

$$\sum_{n=0}^{\infty} c_A(n) = 1 \qquad (20)$$

The function c_A is the basic aging function of the bibliography if growth is not taken into account during the averaging process. Indeed, what we do is looking - over the period under study - at the different citation curves (as distributions). We then take the average number of citations received by an article that is n years old. During this averaging process we do not take into account how many articles have been published in the different years. We will henceforth drop the adjective 'average' because otherwise confusion could arise when taking averages involving the function c_A.

3.3 The global aging distribution function, denoted as c_G.

The function $c_G(n)$ is defined, for $n \in \mathbb{N}$, as:

$$c_G(n) = \sum_{t=0}^{T} j(t) c_t(n) \qquad (21)$$

This is the weighted average of the instantaneous aging distribution functions c_t. Weighting occurs through the functions j(t). Hence it is the aging function of the bibliography weighted for growth over the period [0,T]. It is clear now that in the definition of the aging function $c_A(n)$ all periods were considered with the same weight, or, stated otherwise, weighting was done using a uniform distribution over the interval [0,T]. Note further that also c_G is a bona fide density function, i.e.

$$\sum_{n=0}^{\infty} c_G(n) = 1 \qquad (22)$$

The function c_G is the basic aging distribution function for the bibliography (over a period with length T) when growth is taken into account. Comparison between the functions c_A and c_G (and of derived measures, see further) will reveal important aspects of the influence of growth on aging.

The difference between average and global measures was studied extensively in (Rousseau & van Hooydonk, 1996; Egghe & Rousseau, 1996a,b). From these studies it is known that $c_A(n) < c_G(n)$, $c_A(n) = c_G(n)$ and $c_A(n) > c_G(n)$ are all possible, showing that c_A and c_G have truly different properties.

We recall that, in our opinion, studying the 'influence of growth on obsolescence' over a period of length T, means studying the differences between the aging functions c_A and c_G and functions or measures derived thereof.

3.4 The utilization distribution function c_U

The utilization distribution function $c_U(n)$ is defined, for $n \in \mathbb{N}$, as:

$$c_U(n) = \sum_{t=0}^{\min(n,T)} j(t) c_t(n-t) \qquad (23)$$

It is not difficult to show that also c_U is a distribution function, i.e.

$$\sum_{n=0}^{\infty} c_U(n) = 1 \qquad (24)$$

The function $c_U(n)$ gives the relative number of references to articles that are n years old (as measured from the present, i.e. time 0). Indeed, a reference to a publication that is (now) n years old comes from a reference in a paper that is (now) t years old when at the time of the publication of this paper the cited publication was n-t years old. Here $t \in [0,T]$. The function $c_U(n)$ measures the relative use that has been made of publications published n years ago. This is done independently from the exact time of citation. In the same way as the function j measures the (reversed) growth of articles, the function c_U measures the use that has been made of these articles. In this sense it can be considered as a citation analog of the reversed growth function j.

3.5 The average utilization function c_{AU}

If we replace in (23) the function c_t by c_A we obtain what we call the average utilization function, denoted as c_{AU}:

$$c_{AU}(n) = \sum_{t=0}^{\min(n,T)} j(t) c_A(n-t) \qquad (25)$$

It measures analogous features as c_U, but uses, at every moment, the average aging distribution c_A instead of the instantaneous one. Therefore, also c_{AU} can be considered a citation analog of the reversed growth function j.

From (25) it follows that

$$c_{AU}(t) = (j \star c_A)(t) = (c_A \star j)(t) \qquad (26)$$

where $\star$ denotes the convolution operator. For more information on the convolution operator and its use in informetrics we refer the reader to the tutorial (Rousseau, 1998).

The average utilization function was studied in (Egghe, 1993) and in (Egghe et al., 1995). Yet, in the first one $c_{AU}(t)$ was considered more as an aging function. The more appropriate approach, namely considering it as an utilization function, was taken in the second article.

Density functions such as (23) and (25) really represent the relative use of publications (as gauged by citation counts), considered from the present on (and not from the time the reference was given). This means that, if we consider a scientific domain in which all activity has stopped, say p years ago, and if c_U denotes the density function (23) and $c_{U,p}$ denotes the density function as it was p years ago , i.e. at the time all activity stopped, then, for all $t \geq p$,

$$c_U(t) = c_{U,p}(t-p) \qquad (27)$$

Proof. Put $j_p(t) = j(t+p)$ (and note that $j(m) = 0$ for $0 \leq m < p$). Similarly, we put $c_{m,p}(n) = c_{m+p}(n)$. Then:

$$c_U(t) = \sum_{m=0}^{\min(t,T)} j(m)c_m(t-m) = \sum_{m=p}^{\min(t,T)} j(m)c_m(t-m)$$

Now, putting m = n+p gives:

$$c_U(t) = \sum_{n=0}^{\min(t-p,T-p)} j(n+p)c_{n+p}(t-n-p) = \sum_{n=0}^{\min(t-p,T-p)} j_p(n)c_{n,p}(t-n-p) = c_{U,p}(t-p) \qquad (28)$$

This proves Equation (27).

The previous result illustrates the use of the utilization function. A similar result can be shown for c_{AU}.

Having introduced the basic distribution functions, we now define some derived measures.

4. Derived measures

In the previous sections we showed that growth had a close relation to aging. Indeed, from a methodological point of view, we know that they are both defined and studied in analogous ways, cf. Equations (4),(8),(9). But further, we note that growth has an influence on aging in at least two ways.

The more a field grows, the more articles come into existence, acting as sources for references to the past, i.e. to articles published earlier.
The faster a field grows the heavier the competition between 'older' articles to get into the reference list of the new ones (the dilution effect).

These two opposing dynamic forces act on the aging function of a field. How this actually happens will be shown further on by using the distribution functions c_A and c_G.

Growth data themselves yield 'aging' measures, more specifically on the age of a field.

In this context we define the mean age of a field that started T years ago (measured through publications) as:

$$E(j) = \frac{1}{T+1}\sum_{t=0}^{T} t\, j(t) \qquad (29)$$

Note that

$$E(g) = \frac{1}{T+1}\sum_{t=0}^{T} t\, g(t) = T - E(j) \qquad (30)$$

The median age of the field (or the publication half-life of the field) is denoted as Md(j). The preceding formulae show how age characteristics of a field (mean, median) are defined through the reversed growth function j(t) (or equivalently, the growth function g(t), as Md(g) = T – Md(j), if j = 0 for t > T). Aging on the other hand is defined using the functions c_A or c_G. Here use of the sources, as measured though their productive capabilities (e.g. citations), is of importance.

The aging rate is defined, for all n ∈ ℕ, as:

$$R_A(n) = \frac{c_A(n+1)}{c_A(n)} = \frac{\sum_{t=0}^{T} c_t(n+1)}{\sum_{t=0}^{T} c_t(n)} \qquad (31)$$

This is the aging rate of the field (measured via our bibliography and in such a way that growth has not been accounted for; it is based on observed data). The median of c_A, denoted as Md(c_A), is called the citation half-life. It is the median time between the date of publication and the date of its references if growth is not taken into account. Hence it is also the observed half-life of the field. Finally, the mean citation age, denoted as E[c_A], is defined as:

$$E[c_A] = \sum_{n=0}^{\infty} n\, c_A(n) \qquad (32)$$

It is the average time between the date of a publication and the date of its references, if growth is not taken into account.

Similarly we define the global aging rate, denoted as $R_G(n)$,

$$R_G(n) = \frac{c_G(n+1)}{c_G(n)} = \frac{\sum_{t=0}^{T} j(t)\, c_t(n+1)}{\sum_{t=0}^{T} j(t)\, c_t(n)} \qquad (33)$$

the global citation half-life $Md(c_G)$, and the mean global citation age $E[c_G]$, (we do not repeat the defining formulae here) by replacing the function c_A by the function c_G. The importance of introducing these rates can be appreciated from the following observations. If, for a certain n, $R_G(n) > R_A(n)$ this means that the aging rate n periods ago, is larger if growth is taken into account, than when it is not.

The utilization functions c_U and c_{AU}, being 'citation analogs' of the growth function g, also give information on the age of a discipline, but now from the utilization point of view. We call

$$E[c_U] = \sum_{n=0}^{\infty} n\, c_U(n) \qquad (34)$$

the mean utilization age of the field and $Md(c_U)$ the utilization half-life. Similar definitions can be given with c_{AU} instead of c_U. Also utilization rates can be defined. We think, however, that these are of less importance. Note that, because of (27):

$$E[c_U] = E[c_{U,p}] + p \qquad (35)$$

This is an obvious result. Its easy proof is left to the reader. Property (35) illustrates that the naming 'mean utilization age' is appropriate.

5. Properties

5.1 Aging

When studying a whole period of length T, the influence of growth on aging (obsolescence) can best be studied by comparing the aging rates $R_A(n)$ and $R_G(n)$. The next theorem gives the basic result.

Theorem 1

Supposing the growth function g to be strictly increasing on [0,T], i.e. $\alpha(t) > 1$, the following assertions are true:

a) If the instantaneous aging rate functions $\beta_t(n)$ are strictly increasing in the variable $t \in [0,T]$, i.e. $\forall\, t \in [0,T]$ and $\forall\, n \in \mathbb{N}$:

$$\frac{c_t(n+1)}{c_t(n)} < \frac{c_{t+1}(n+1)}{c_{t+1}(n)} \qquad (36)$$

then, for every $n \in \mathbb{N}$:

$$R_A(n) > R_G(n) \qquad (37)$$

In words: under these circumstances an increasing growth reinforces aging.

b) If the instantaneous aging rate functions $\beta_t(n)$ are strictly decreasing in the variable $t \in [0,T]$, then, for every $n \in \mathbb{N}$:

$$R_A(n) < R_G(n) \qquad (38)$$

In words: under these circumstances aging diminishes, although growth increases.

c) If the instantaneous aging rate functions $\beta_t(n)$ are constant in the variable $t \in [0,T]$, then, for every $n \in \mathbb{N}$:

$$R_A(n) = R_G(n) \qquad (39)$$

In words: under these circumstances an increasing growth has no influence on aging.

Proof of a): We have to show that for every $n \in \mathbb{N}$:

$$\frac{\sum_{i=0}^{T} c_i(n+1)}{\sum_{s=0}^{T} c_s(n)} > \frac{\sum_{i=0}^{T} j(i)c_i(n+1)}{\sum_{s=0}^{T} j(s)c_s(n)} \qquad (40)$$

or, equivalently that:

$$\sum_{i,s} c_i(n+1)c_s(n)(j(s)-j(i)) > 0 \qquad (41)$$

From (36) we know that, for every $n \in \mathbb{N}$, and for every i and s such that $0 \leq i < s \leq T$,

$$\frac{c_i(n+1)}{c_i(n)} < \frac{c_s(n+1)}{c_s(n)} \qquad (42)$$

or: $c_s(n+1)c_i(n) - c_s(n)c_i(n+1) > 0$. We know, moreover, that for $0 \leq i < s \leq T$, $j(i) - j(s) > 0$ (g increases, hence j decreases). Now, Equation (41) can be rewritten as:

$$\sum_{s=2}^{T}\sum_{i=1}^{s-1}\left(c_s(n+1)c_i(n) - c_i(n+1)c_s(n)\right)\left(j(i)-j(s)\right) > 0$$

This inequality is clearly satisfied as each factor and each term of the left-hand side is positive.

The cases b) and c) are similar and are left to the reader.

Remark 1. The above result shows that *an increasing growth can influence aging but that it does not cause aging*. This is most obvious from Equation (39).

Remark 2. If g(t) is constant trivially $R_A = R_G$. If g(t) is decreasing we obtain the opposite results of those obtained in Theorem 1.

Remark 3. It follows from the proof that it suffices that g increases strictly at least once in [0,T] to obtain the results of Theorem 1.

In the next section (6.2) we offer examples of distribution functions that actually meet the requirements of Theorem 1.

5.2 Expected citation age and regression lines

The influence of growth on citation age can be studied by comparing the averages $E[c_A]$ and $E[c_G]$. We first note that:

$$E[c_A] = \sum_{n=0}^{\infty} n c_A(n) = \sum_{n=0}^{\infty} n \left(\frac{1}{T+1} \sum_{t=0}^{T} c_t(n) \right) = \frac{1}{T+1} \sum_{t=0}^{T} E[c_t] \qquad (43)$$

and

$$E[c_G] = \sum_{n=0}^{\infty} n c_G(n) = \sum_{n=0}^{\infty} n \left(\sum_{t=0}^{T} j(t) c_t(n) \right) = \sum_{t=0}^{T} j(t) E[c_t] \qquad (44)$$

We see that $E[c_A]$ is the average of the numbers $E[c_t]$, t = 0,...,T and that $E[c_G]$ is the weighted average of the $E[c_t]$, t = 0,...,T, weighted by j(0),j(1),...,j(T). The numbers $E[c_t]$ themselves refer to the mean age of the references of a document published t years ago. In order to prove a result relating these two quantities we recall a theorem shown in (Egghe & Rousseau, 1996a).

Theorem 2 (Egghe & Rousseau, 1996a, Theorem 2)

If $(x_i)_i$, i = 1,...,N is a set of positive numbers, and $f(x_i) = y_i/x_i$, i = 1,..,N, where also the $(y_i)_i$ are positive numbers, and if r denotes the slope of the regression line of f(x) over x, then

$$r > 0 \Leftrightarrow \frac{\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} x_i} > \frac{1}{N}\sum_{i=1}^{N}\frac{y_i}{x_i} \qquad (45)$$

and, moreover,

$$r = 0 \Leftrightarrow \frac{\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} x_i} = \frac{1}{N}\sum_{i=1}^{N}\frac{y_i}{x_i} \qquad (46)$$

This leads to the following theorem on the influence of growth on age.

Theorem 3

$E[c_G] > E[c_A]$ if and only if the slope of the regression line of the points $((j(0),E[c_0]),\dots (j(T), E[c_T]))$ is positive. Moreover, if g, and hence j, are not constant functions, $E[c_G] = E[c_A]$ if and only if the slope of this regression line is zero.

Proof.

We apply Theorem 2 with j in the role of x and j E[c] in the role of y. We obtain from (45) that the slope of this regression line is strictly positive if and only if

$$\frac{\sum_{t=0}^{T} j(t)E[c_t]}{\sum_{t=0}^{T} j(t)} > \frac{1}{T+1}\sum_{t=0}^{T} E[c_t] \quad or \quad \sum_{t=0}^{T} j(t)E[c_t] > \frac{1}{T+1}\sum_{t=0}^{T} E[c_t] \quad or \quad E[c_G] > E[c_A]$$

where we have used Equations (43) and (44).

The other assertion results similarly from the other part of Theorem 2.

Theorem 3 states that $E[c_G] > E[c_A]$ if and only if $E[c_t]$ grows (on average) as the reversed growth function j(t). In other words: $E[c_G] > E[c_A]$ if and only if – on the whole – $E[c_t]$ is large whenever j(t) is.

Note the fact that we have obtained a characterization, i.e. a necessary and sufficient condition, for the relation between $E[c_A]$ and $E[c_G]$.

Note also that, as in the previous section, growth alone does not cause aging nor a decrease or increase of the average citation age. Moreover, if the yearly growth is constant, clearly $E[c_G] = E[c_A]$.

Concerning the citation half-lives (medians of c_A and c_G) we obtain a similar result as follows. We first note that, for every $m \in \mathbb{N}$:

$$\sum_{n=0}^{m} c_A(n) = \frac{1}{T}\sum_{t=0}^{T}\sum_{n=0}^{m} c_t(n) \qquad (47)$$

and

$$\sum_{n=0}^{m} c_G(n) = \sum_{t=0}^{T} j(t) \sum_{n=0}^{m} c_t(n) \qquad (48)$$

If now

$$\sum_{n=0}^{m} c_A(n) < \sum_{n=0}^{m} c_G(n) \qquad (49)$$

holds for every m, then clearly:

$$\mathrm{Md}(c_A) > \mathrm{Md}(c_G) \qquad (50)$$

A concrete example is given in the following proposition. First, we introduce the notation:

$$C_t(m) = \sum_{n=0}^{m} c_t(n) \qquad (51)$$

Proposition 4

If for every $m \in \mathbb{N}$, $C_t(m)$ is strictly increasing whenever j(t) increases then

$$\mathrm{Md}(c_A) > \mathrm{Md}(c_G) \qquad (52)$$

Proof. If for every $m \in \mathbb{N}$, $(j(t), C_t(m))$, t = 0,...,T increases, its regression line also increases (Egghe & Rousseau, 1996a). Hence, applying Theorem 2 with j in the role of x, and $j(t)\, C_t(m)$ in the role of y, we obtain:

$$\frac{\sum_{t=0}^{T} j(t) C_t(m)}{\sum_{t=0}^{T} j(t)} > \frac{1}{T+1} \sum_{t=0}^{T} C_t(m) \quad or \quad \sum_{n=0}^{m} c_G(m) > \sum_{n=0}^{m} c_A(n) \qquad (53)$$

From this, it follows (by the definition of a median) that $\mathrm{Md}(c_A) > \mathrm{Md}(c_G)$.

Although the proofs of Theorem 3 and Proposition 4 are similar, their results seem contradictory. Note, however, that the condition that $C_t(m)$ is strictly increasing in j(t), for every m $\in \mathbb{N}$ implies that the numbers $c_t(n)$ must be decreasing in j(t), (keeping in mind that

$$\sum_{n=0}^{\infty} c_t(n) = 1$$

for all t in [0,T]), which explains the conclusion.

6. Examples

6.1 Exponential growth and exponential aging

Although other functions can be used (see further) the exponential functions are the basic functions to describe growth and aging. Consequently, we first show what the basic measures are in the exponential case.

For the ease of calculation we assume that all c_t-functions have the same form: for every n $\in \mathbb{N}$, and every t $\in$ [0,T],

$$c_t(n) = k\,c^n \qquad (54)$$

As the c_t-functions are distributions, k = 1 – c, with c < 1. Similarly, we define the growth function g(t) for t $\in$ [0,T], (and do not forget that the value 0 for the function g is T time periods ago, while the value T for g is 'now') as:

$$g(t) = K\,g^t \qquad (55)$$

Because also g is a discrete distribution function (but now on the interval [0,T]), we have: K = (g-1)/(g^{T+1} – 1) for g $\neq$ 1. If g = 1, the growth function is constant, and K

= 1/(1+T). Note also that if $g > 1$ the growth function increases (the most natural case), while if $0 < g < 1$, the growth function decreases. We further assume that $g \neq c$. The reversed growth function j(t), $t \in [0,T]$, is equal to $g(T-t) = K\, g^{T-t} = L\, g^{-t}$, with $L = Kg^T$.

The growth rate is here constant, namely equal to g, as are the instantaneous aging rates: $\beta_t = c$. Hence the aging distributions $c_A(n)$ and $c_G(n)$ are both equal to $k\, c^n$. Further, as to utilization, we obtain (Egghe et al. 1995):

$$c_U(n) = c_{AU}(n) = \sum_{t=0}^{\min(n,T)} Kg^{T-t}\, k\, c^{n-t} = Kkg^T c^n \frac{1-(gc)^{-(n+1)}}{1-(gc)^{-1}},\ if\ n<T$$
$$= Kkg^T c^n \frac{1-(gc)^{-(T+1)}}{1-(gc)^{-1}},\ if\ n \geq T \qquad (56)$$

Note that utilization is, logically, a mix between c and g. We now proceed with the calculation of derived measures. The (average) age of articles that belong to an exponentially growing field, is:

$$E[j] = \sum_{t=0}^{T} t\, j(t) = L \sum_{t=0}^{T} t\, g^{-t} = L\, g \frac{(g^{-2}) - (T+1)g^{-(T+2)} + T\, g^{-(T+3)}}{(1-g^{-1})^2}$$

$$= \frac{g^{T+1} - (T+1)g + T}{(g^{T+1}-1)(g-1)} \qquad (57)$$

As example we take a field that annually grows by 10%, so g = 1.1, and began 20 years ago (T = 20), then the average age of an article of this field is 6.7 years. For this example the median age of an article in this field is 20 – Md(g) = 20 – 16.05 = 5.95 year.

We clearly have: $E[c_A(n)] = E[c_G(n)] = c/(1-c)$. If $c = 0.8$ the average age of a reference is 4 years. Of course, all rates of c_A or c_G are equal to c. Further: $E[c_U] = E[c_{AU}] =$

$$\frac{Kkg^T}{1-(gc)^{-1}}\left(\frac{c}{(1-c)^2}-\frac{g^{-1}-(T+1)g^{-(T+1)}+Tg^{-(T+2)}}{cg(1-g^{-1})^2}-\frac{(T+1)c^{T+1}-Tc^{T+2}}{(gc)^{T+1}(1-c)^2}\right)$$

For the rate functions of c_U (= c_{AU}) we obtain (for $n < T$):

$$R_U(n)=\frac{c_U(n+1)}{c_U(n)}=\frac{c\left(1-(gc)^{-(n+2)}\right)}{1-(gc)^{-(n+1)}} \qquad (58)$$

For n large (but smaller than T) this expression is approximately 1/g, hence smaller than 1 if $g > 1$, but larger than 1 if $g < 1$. This means that for such n the utilization rate may increase or decrease depending on the growth rate. However, for the most common case that $g > 1$ the utilization rate decreases. For $n > T$ it is equal to c, hence smaller than 1 (the utilization rate decreases). More information and results on utilization functions in this context may be found in (Egghe et al. 1995).

6.2 An illustration of Theorem 1

We next show that it is possible to have citation distributions that satisfy the requirements of Theorem 1.

Let (α_t) be a sequence of strictly positive numbers, then we put, for all $t \in T$:

$$c_t(n)=\left(1-e^{-\alpha_t}\right)e^{-\alpha_t n} \qquad (59)$$

It is easy to verify that for every t, c_t is a distribution function, i.e. all $c_t(n)$ are non-negative numbers with sum (for n = 0, 1, ...) equal to 1. Now,

$$\beta_t(n) = e^{-\alpha_t} \qquad (60)$$

Consequently, $\beta_t(n)$ is increasing in t if the sequence α_t is decreasing (take e.g. α_t = 1/t); similarly, if α_t is increasing (take e.g. α_t = t), then $\beta_t(n)$ is decreasing in t. Of course, a constant sequence α_t leads to a sequence $\beta_t(n)$ that is constant in t. Note that the citation distribution is in any case a decreasing exponential function in n.

6.3 Even if j(t) and all $c_t(n)$ are decreasing (respectively in t and in n), which means that the bibliography shows an increasing growth, it is possible that the relative impact rate, which is equal to the instantaneous corrected aging rate, is increasing in time.

Example: let the reversed growth function be $j(t) = A/t^3$; and let $c_t(n) = B_t/((n+1)^2t)$, then clearly j and the aging functions $c_t(n)$ are strictly decreasing. The rate functions ρ_t are (by (9)):

$$\rho_t(n) = \frac{B_t}{(n+2)^2 t}\frac{(n+1+t^3)}{A}\frac{A}{(n+t)^3}\frac{(n+1)^2 t}{B_t} = \frac{(n+1)^2}{(n+2)^2}\frac{(n+1+t)^3}{(n+t)^3} \qquad (61)$$

For t = 0 (now!) this function is decreasing in n but always larger than 1. This means that, relatively speaking (this means: taking growth into account), the older literature has a larger impact than the more recent one. For larger t this rate increases and, again, becomes larger than one for large n (although not for small values of n).

7. Previous work and its relation to ours

The "half-life" concept was made popular in the information sciences by Burton and Kebler (1960). It is used as a summary statistic to describe the rate of obsolescence of a scientific literature. They define it as *the time during which one-half of all currently active literature was published.* It is clear that this is a synchronous definition. 'Active' can mean 'being cited', but also 'being borrowed', etc... In this article we have used the term 'citation half-life'. They calculated half-life values for the periodical literature (excluding books) for different fields, based on a fitted curve of the form:

$$y(t) = 1 - \left(\frac{a}{e^t} + \frac{b}{e^{2t}} \right) \qquad (62)$$

where a + b =1, and y denotes the cumulated fraction of cited literature (values between 0 and 1); t denotes time. They conclude that a short half-life, i.e. rapid obsolescence, is the result of rapidly changing techniques or interest. To some extent, rapid obsolescence could also be the result of poor quality. Note, though, that Equation (62) is only the result of a fitting exercise. No meaning should be given to the two exponential functions separately (Motylev, 1989).

MacRae Jr.(1969), following Price (1965), assumes an exponential growth of the literature combined with an exponential time-independent corrected citation distribution (our terminology). He comes to the conclusion that the observed citation distribution must be exponential too. This conclusion is verified (and shown to be correct) for data taken from the Physical Review (1957), the biomedical sciences and the American Sociological Review.

Line (1970) states that more rapidly the literature is growing, the shorter the half-life of a literature, unless the number of citations per article is decreasing at the same time. The so-called 'half-life' of a literature is therefore composed of its obsolescence rate and its growth rate. All this is quite correct! Taking all rate functions as constants, he defines the apparent obsolescence factor A as c_{n-1}/c_n (hence his A is equal to our $1/\beta_t(n-1)$ or $1/\beta$ if we consider this rate also as a constant). Line continues by introducing c_n/G, where G is the growth factor (his terminology). In our notation and terminology G is the growth rate, α, and is equal to $1/\chi$ (in the case of constant rates). Finally, Line introduces the corrected obsolescence factor as D = A.G . Here his D is defined as $c_{n-1}/(c_n/G)$. In our notation, and with variable rates this becomes:

$$D(n)=\frac{c_t(n-1)}{c_t(n)}\frac{1}{\chi(n-1)}$$

or, with constant rates:

$$D=\frac{1}{\beta\chi}$$

Further, Line derives the corrected half-life, denoted as h, as a function of A and G; and as a function of G and the observed half-life (denoted as m).

$$h=\frac{\ln(0.5)}{\ln(A)+\ln(G)} \qquad (63)$$

or, as a function of the observed half-life:

$$\frac{1}{h} = \frac{1}{m} - \frac{\ln(G)}{\ln(2)} \qquad (64)$$

Note though that Line forgets to mention that his expression for the corrected half-life is only valid for D < 1. Otherwise it leads to invalid results. If he took e.g. G = 1.1 and A = 0.98, he would find a corrected half-life of –9.2 (?!). In an added editorial note (Vickery, 1970) a referee remarks that D is the rate at which the likelihood of use declines with age.

In a recent article in Nature Abt (1998) poses the question ' Why have some papers long citation lifetimes?' He considers the special case of the Astrophysical Journal and Physical Review. He finds that the speed of growth of the field is a crucial factor and proposes 'half-lives corrected for growth'. Although Abt does a diachronous study (with 1954 as starting date for articles in the Astrophysical Journal) and provides very little real information it seems that he has assumed a linear growth of the literature. Indeed, he writes that the Astrophysical Journal has an observed citation half-life of 29.3 years and that the number of articles has grown over 40 years by a factor of 11. From a citation half-life of 29.3 years we can derive a citation function of the form e^{-at}, with a = 0.024. Assuming a linear growth of the literature and an exponential form of the 'corrected' citation function yields $e^{-at} = ct\, e^{-bt}$ with 40c = 11. This leads to b = 0.084 and hence a 'corrected' half-life of approximately 8.3 years. Abt gives as 'corrected' value 8.0, so that we assume that our reasoning is right. The main difference between Line's ideas and Abt's ones is the fact that Line assumes an exponential growth of the literature while Abt (probably) assumes linear growth. We think that both assumptions are arguable: Line considers the whole literature (which is often assumed to grow exponentially, cf. Price (1965), Van Raan (2000)), while Abt considers a single journal. The number of articles in a single journal never grows exponentially. Note though that Abt seems to assume that the growth rate of the Astrophysical Journal is characteristic for the field as a whole.

Wallace (1986), following an idea of Buckland (1972), studied the following hypothesis: 'For a given subject literature, the median citation ages of the journals contributing to that literature will vary inversely with the productivity of those journals, where productivity is measured in terms of the number of articles contributed by each journal'. This question cannot be answered or studied within the framework of this article. Other models must be developed to do this. By the way, Wallace did not find a general confirmation of his hypothesis (nor did Buckland).

8. Conclusions

We have shown how to correct an observed citation distribution (c_t) for growth, once the growth distribution is known. This is done in the spirit of Line's earlier work. We have, moreover, shown that this corrected aging distribution coincides with the distribution of yearly impacts (citations per publication).

More interestingly, we have shown how the influence of growth on aging can be studied over a complete period as a whole. Here the difference between the so-called average (c_A) and global (c_G) aging distributions is the main factor. Also utilization functions, describing the use of sources over a specific period, are defined. Our main results are that growth can influence aging but that it does not cause aging. Further, using regression techniques, we have shown the following relation between expectations: $E[c_G] > E[c_A]$ holds, if and only if $E[c_t]$ is large whenever the reversed growth function is.

Examples of the use of these functions and techniques are given. Finally, we have placed some classical articles on aging and obsolescence in our framework.

Acknowledgments. The authors thank the anonymous referees for stimulating observations, and M. Dekeyser (KHBO –Ostend) for help in the preparation of this article.

References

Abt, H.A. (1998). Why some papers have long citation lifetimes. Nature, 395, 756-757.

Brookes, B.C. (1970). Obsolescence of special library periodicals: sampling errors and utility contours. Journal of the American Society for Information Science, 21, 320-329.

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

Burton, R.E. and Kebler, R.W. (1960). The "half-life" of some scientific and technical literatures. American Documentation, 11, 18-22.

Clark, C.V. (1976). Obsolescence of the patent literature. Journal of Documentation, 32(1), 32-52.

De Groote, S., Rousseau, R. and Vervliet, H.D.L. (1991). The burgundy effect in off-line union catalogs of serials. Library and Information Science Research, 13(1), 3-19.

Diodato, V. and Smith, F. (1993). Obsolescence of music literature. Journal of the American Society for Information Science, 44(2), 101-112.

Egghe, L. (1988). Mathematical relations between impact factors and average number of citations. Information Processing & Management, 24, 567-576.

Egghe, L. (1993). On the influence of growth on obsolescence. Scientometrics, 27, 195-214.

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

Egghe, L. and Rao, I.K. Ravichandra (1992b). Classification of growth models based on growth rates and its applications. Scientometrics, 25, 5-46.

Egghe, L., Rao, I.K. Ravichandra and Rousseau, R. (1995). On the influence of production on utilization functions: obsolescence or increased use? Scientometrics, 34, 285-315.

Egghe, L. and Rousseau, R. (1990). Introduction to Informetrics. Amsterdam: Elsevier.

Egghe, L. and Rousseau, R. (1996a). Average and global impact of a set of journals. Scientometrics, 36, 97-107.

Egghe, L. and Rousseau, R. (1996b). Averaging and globalizing of informetric and scientometric data. Journal of Information Science, 22(3), 165-170.

Griffith, B.C., Servi, P.N., Anker, A.L. and Drott, M.C. (1979). The aging of scientific literature: a citation analysis. Journal of Documentation, 35(3), 179-196.

Line, M.B. (1970). The 'half-life' of periodical literature: apparent and real obsolescence. Journal of Documentation, 26, 46-52.

Matricciani, E. (1991). The probability distribution of the age of references in engineering papers. IEEE Transactions on Professional Communication, 34(1), 1991, 7-12.

McRae Jr., D. (1969). Growth and decay curves in scientific citations. American Sociological Review, 34, 631-635.

Motylev, V.M. (1989). The main problems of studying literature aging. Scientometrics, 15, 97-109.

Nakamoto, H. (1988). Synchronous and diachronous citation distributions. In: Informetrics 87/88 (Egghe, L. & Rousseau, R. eds.). Amsterdam: Elsevier 157-163.

Narayana, N.S.S. (1998). Random reflections on some aspects of informetrics. In: DRTC Workshop on Informetrics and Scientometrics (Bangalore) (I.K. Ravichandra

Rao, ed.). Bangalore: Indian Statistical Institute Documentation Research and Training Centre, Paper BA.

Pierou, V., Nunes Amaral, L.A., Gopikrishnan, P., Meyer, M. and Stanley, H.E. (1999). Similarities between the growth dynamics of university research and of competitive economic activities. Nature, 400, 433-437.

Price, D. J. de Solla (1965). Networks of scientific papers. Science 149, 510-515.

Rousseau, R. (1988). Citation distribution of pure mathematics journals. In: Informetrics 87/88 (Egghe, L. & Rousseau, R. eds.). Amsterdam: Elsevier, 249-262.

Rousseau, R. (1997). Sitations: an exploratory study. Cybermetrics, 1. **http://www.cindoc.csic.es/cybermetrics/articles/v1i1p1.html**

Rousseau, R. and Van Hooydonk, G. (1996). Journal production and journal impact factors. Journal of the American Society of Information Science, 47(10), 775-780.

Rowlands, I. (1999). Patterns of scholarly communication in information policy: a bibliometric study. Libri, 49, 59-70.

Scales, P. (1976). Citation analyses as indicators of the use of serials: a comparison of ranked title lists produced by citation counting and from use data. Journal of Documentation, 32(1), 17-25.

Stinson, E.R. (1981). Diachronous vs. synchronous study of obsolescence. PhD dissertation, University of Illinois (UMI #81-27705).

Stinson, E.R. and Lancaster, F. (1987). Synchronous versus diachronous methods in the measurement of obsolescence by citation studies. Journal of Information Science, 13, 65-74.

Tague, J., Beheshti, J. and Rees-Potter, L. (1981). The law of exponential growth: evidence, implications and forecasts. Library Trends, 30(1), 125-149.

Tsay, M.-Y. (1998). Library journal use and citation half-life in medical science. Journal of the American Society for Information Science, 49(14), 1283-1292.

Van Raan, A.F.J. (2000). On growth, aging, and fractal differentiation of science. Scientometrics (to appear).

Vickery, B.C. (1970). Editorial note. Journal of Documentation, 26(1), 53-54.

Wallace, D.P. (1986). The relationship between journal productivity and obsolescence. Journal of the American Society for Information Science, 37, 136-145.