Jointly published by Kluwer Academic Publishers, Dordrecht and Akadémiai Kiadó, Budapest

Scientometrics, Vol. 47, No. 2 (2000) 347–362

On growth, ageing, and fractal differentiation of science

ANTHONY F. J. VAN RAAN

Centre for Science and Technology Studies, University of Leiden, Leiden (The Netherlands)

On the basis of the measured time-dependent distribution of references in recent scientific publications, we formulate a novel model on the ageing of recent scientific literature. The framework of this model is given by a basic set of mathematical expressions that allows us to understand and describe large-scale growth and ageing processes in science over a long period of time. In addition, a further and striking consequence results in a self- consistent way from our model. After the Scientific Revolution in 16th century Europe, the 'Scientific Evolution' begins, and the driving processes growth and ageing unavoidably lead – just as in our biological evolution – to a fractal differentiation of science. A fractal structure means a system build up with sub-systems characterised by a power-law size distribution. Such a distribution implies that there is no preference of size or scale. Often this phenomenon is regarded as a fingerprint of self-organisation. These findings are in agreement with earlier empirical findings concerning the clustering of scientific literature. Our observations reinforce the idea of science as a complex, largely self-organising 'cognitive eco-system'. They also refute Kuhn's paradigm model of scientific development.

Introduction

Almost 35 years ago *de Solla Price* published the distribution of the publication years of references given in 1961 papers covered by the *Science Citation Index*. The references (or: cited documents) spanned a period of hundred years: 1862 through 1961.[1–4] Several remarkable observations were made. We mention the two most striking features. First, the *exponential* increase of the number of cited documents. Second, the two dips marking the periods of World Wars I and II, immediately followed by an astonishing 'recovery' of science after the wars toward a level more or less extrapolated from the period *before* the wars.

The crucial element in the analysis of such *age distributions* of references of a given 'citing' year is the interpretation of the findings in terms of *growth and ageing* of science. Is it 'just' the age-distribution of references? Clearly this cannot be the case: although undoubtedly ageing of earlier published work is part of reality, another part of reality is that in earlier times there were much less documents published than in recent

0138–9130/2000/US $ 15.00

times. Thus, the time-dependent distribution of references will always be a specific combination of the ageing and growth phenomena of science, or better said: of scientific literature. After the publication of Price's age-distribution of references, this problem was intensively discussed.[5,6]

Furthermore, the use of any database to observe whatever processes, is in fact methodologically similar to the use of observation instruments in, for instance, physics or astronomy. This means that characteristics of the instruments (in this case: the database) will also be reflected in the measurements. Even worse, characteristics of databases may change in the course of time. However, in a first and good approximation one may state that the *Science Citation Index* represents quite well a major, and generally important part of the scientific literature. Second, by using one fixed publication year T, age-distributions of cited documents (the references of that year T) will not reflect time-dependent database characteristics in, for instance, the registration of references.

On the basis of empirical data, a new hypothesis on literature ageing, and assumptions on the relation between growth and number of new developments, we arrive at a model for growth and differentiation of science. The empirical data is essentially the age-distribution of references as given in publications of last year, i.e., T = 1998. Therefore, our approach can be characterised as a 'synchronous' study (time-dependence measurement based on one fixed 'source' year). In a forthcoming paper[7] we discuss this matter in more detail and, in particular, 'diachronous' aspects (i.e., comparison of cited document time-dependence originating from different 'source' years).

Empirical data: first observations and questions

Using our advanced bibliometric data-system as described in Refs 8 and 9, we identified all references (order of magnitude: 15,000,000) of all papers published in 1998 and covered by the *Science Citation Index* and its sister-indexes (total number of publications about 1,000,000). Next, we sorted these references by their year of publication t (thus $t \leq T = 1998$) and counted the number of references from the year 1500 up till 1998. As we will show in a forthcoming paper, the period between 1500 and 1800 is characterised by very low numbers and therefore quite 'noisy'. Basically, this figure represents our source of empirical data. We will confine ourselves here to the period from 1800 on. The result of this analysis is given in Fig. 1. We make the following observations together with the formulation of central questions that will be tackled in later sections.

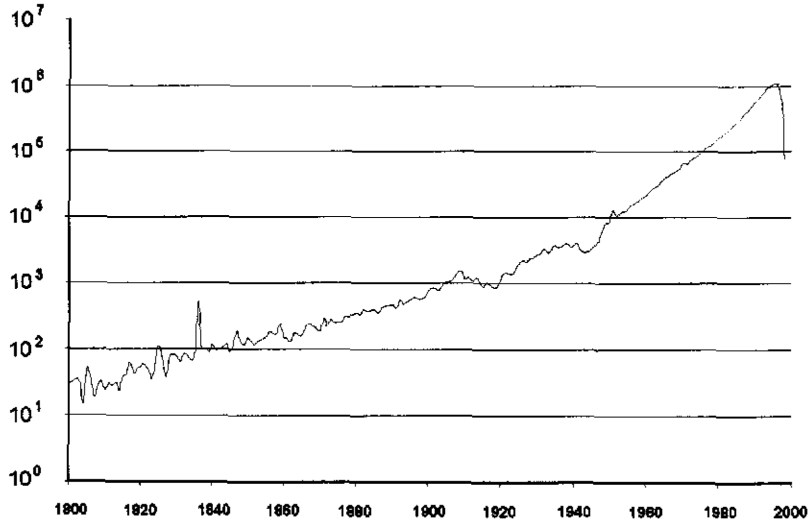

Fig. 1. Age-distribution of all references given in publications of 1998

1. At first sight, our findings fit in with *Price*'s[2] analysis. But a more detailed inspection reveals remarkable differences. We here mention the two most striking. First, of course, our analysis now fully shows the post-war period up to almost the end of this century, whereas *Price*'s analysis stops in a 'recovery' period relatively short after World War II. The importance of this extension is immediately evident: the exponential time-dependence (note that the ordinate of Fig. 1 is logarithmic) appears to gear up to a considerably larger growth-coefficient, resulting in a steeper slope in the figure for the period after 1950.

This observation leads to our first question. Has a rather drastic change of the science system, notably by stronger interaction with the economic and technological environment after the World War II led to a stronger growth? Or should the dynamics of the science system as measured over a much longer period such as in our case, two centuries, not be described by an exponential growth, but by a 'super-exponential growth'?

Our second observation concerns the interbellum-period 1920–1935. In *Price*'s analysis this period shows a 'hill' when extrapolating the exponential increase of the pre-World War I period. This 'hill' is almost disappeared in our analysis. Indeed, if we analyse age-distributions of earlier publication years between 1980 and 1998, we notice a gradual fading away of this 'hill'. In the forthcoming paper[7] we discuss this phenomenon more in detail.

2. From about 1984, we observe a final increase of 'growth', followed by a peak in 1995 and a decrease in the last few years, up to 1998. This observation is very essential in the development of our model. The latter phenomenon, the decrease in the last few years just before 1998, can easily be explained. It is well known, particularly in the natural science and medical fields, that publications of a given year (here: 1998) have a peak in the distribution of their references around the age of three years. Therefore, in the last three years the number of references will decrease. Thus, the crucial question concerns the first phenomenon, the origin of the 'final increase' in the slope of the measured curve before the peak, starting around 1984. This 'final increase' is clearly visible if one fits in Fig. 1 the curve between, say, 1950 and 1980 with a straight line and extrapolates this line up to the end of the curve.

3. In several places in the curve we observe smaller 'peaks', e.g., around 1836, 1951, 1970. The 1951 and 1970 'singularities' in the age-distribution of references mark two extremely highly cited papers. In 1951 it is the publication of O.H. Lowry in the *Journal of Biological Chemistry* on a method to measure the protein percentage particularly for medical diagnosis. The 1970 peak is the *Nature* publication of U.K. Laemmli on a method of molecular cloning. Both papers receive around 10,000 citations each year, which makes them clearly visible, even on the 'meta-level' of our analysis. The 1836 peak is probably an artefact resulting from a remarkable accumulation of US patents around that year. Although patents represent only a small minority in references in the scientific literature, a 'local' accumulation will show up as a peak in the curve. We are currently investigating this phenomenon in more detail.

In the above we formulated questions directly based on our observations. In the next section we tackle these problems with help of a mathematical model.

Mathematical model for growth and ageing

General representation of the system's dynamics

Our first question concerns the search for an explanation, expressed in a mathematical function, for the graph observed in Fig. 1 as a whole. As we already noticed, there are two possible solutions.

First, given the linear relationship over long periods of time between the logarithm of the number of references R(t) and the time t, we describe the time-dependence of the number of references by the differential equation

$$\mathrm{d}\,R(t)/\,\mathrm{d}t = a\,R(t)\,. \tag{1}$$

This is similar to a simple growth process well known in biology. But we remind that an age-distribution of references given in recent papers will reflect two independent processes: *ageing* (scientists will be increasingly less interested in increasingly older literature), and *growth* (there were simply much less papers published in 1930 as there are in 1990, so there is less to be cited to earlier years). We remark that in a more precise approximation, particularly at a lower aggregation level (i.e., on the level of scientific fields instead of science as a whole) a possible interdependence of growth and ageing may have to be taken into account. We return to this point at the end of this section.

Only in the case without ageing, the numbers of references R(t) directly reflect the numbers of papers N(t) in any specific year t, so

$$R(t) = r\,N(t)\,, \tag{2}$$

in which case the age-distribution graph can be interpreted as the growth curve of science, as far as reflected by the number of publications. But as discussed above, we shall have to disentangle the ageing process from the measured curve, before we can analyse the growth of science on the basis of reference age-distributions.

A second possible solution concerns a 'super-exponential' process. This super-exponential approach is inspired by inspecting the age-distribution curve over the whole period of time, 1800–1998, and observing that although the ordinate is logarithmic, there still is a clear non-linear time-dependent increase. This leads to a rather exotic differential equation:

$$\mathrm{d}\,R(t)/\mathrm{d}t = a\,e^{kt}\,R(t)\,. \tag{3a}$$

As far as we could trace up till now, the only dynamical process in the natural sciences described by such a differential equation is the stage of unlimited tumour growth,[10] known as Gompertz growth.[11] (Notice how nice it is to study the age-distribution of references in very recent literature, back to 1800, and having yourself a (necessary!) reference to a paper of 1825!) The Gompertz equation has been applied for the first time in scientometric research on growth phenomena by *Egghe* and *Rao*[12] and by *Rousseau*[13] for citation processes. We will discuss later[7] this particular dynamical process in more detail, but prove here that the Gompertz equation (3a) approaches the 'normal' equation (1) for smaller periods of time. We evaluate the differential equation (3a) as follows:

$$\mathrm{d}\,\mathrm{R}(t)/\mathrm{d}t = a\,e^{kt}\,\mathrm{R}(t) \quad \text{yields}$$

$$[1/\mathrm{R}(t)].\,\mathrm{d}\,\mathrm{R}(t)/\mathrm{d}t = a\,e^{kt}\ , \tag{3b}$$

which leads to

$$\mathrm{d}\,ln\mathrm{R}(t)/\mathrm{d}t = a\,e^{kt}\ , \tag{3c}$$

thus

$$\mathrm{d}\,ln\mathrm{R}(t) = a\,e^{kt}\,\mathrm{d}t\ , \tag{3d}$$

and integrating this equation with limits of integration 0 and t, we have:

$$\int_0^t \mathrm{d}\,ln\mathrm{R}(t) = \int_0^t a\,e^{kt}\ \mathrm{d}t\ , \tag{3e}$$

yielding

$$ln\mathrm{R}(t) - ln\mathrm{R}(0) = (a/k)[e^{kt}-1]\ , \tag{3f}$$

which gives

$$\mathrm{R}(t) = \mathrm{R}(0)\exp\{(a/k)[e^{kt}-1]\}\ , \tag{3g}$$

and with a series expansion of the second exponential part we find for small k and not too large t:

$$\mathrm{R}(t) = \mathrm{R}(0)\exp\{(a/k)[1+kt+\ldots-1]\} \cong \mathrm{R}(0)\,e^{at}\ , \tag{3h}$$

which is the solution of the simple differential equation (1). Thus, although the dynamics of a system as a whole may be characterised by a super-exponential Gompertz equation, it can be approximated for smaller time-tracks by the normal exponential equation (1) with solution (3h). This exercise is necessary for essential further elements of our mathematical model discussed next in this paper.

Solution of the ageing problem

We specify the number of references $R(t)$ a little bit more precisely with $R(T, t)$, where T is the year of the publications from which the references are analysed, in this case $T=1998$.

Now we state the following hypothesis: *the remarkable final increase in the slope of the graph in Fig. 1, concerning the period t = 1984 up to 1995, as discussed previously, reflects the ageing process of recent literature.*

This hypothesis is supported by our further recent observations[7] that this 'final increase' is visible for other (earlier) publication years T, and is always located in a similar position *with respect to T.*

So we have a strong indication that a typical ageing process is particularly active in an exponential fashion (as well known from the age-distribution of individual publications) for a period up to 10, 15 years back from T. For earlier years, it appears that this ageing process is much less strong, so that either 'ageing' is still present but with a much 'slower' exponent, or that, in first approximation, there is no ageing anymore. In the latter case the number of references will be proportional to the number of papers available at that time, a situation described by Eq. (2). This means, that the 'old' literature becomes a relatively 'fixed' archive.

The above model can be described mathematically in the following way.

Let us denote the year where the 'final increase' starts by T_0, and the year of the peak-value by T_1. Then we may write for the number of references $R(T, t)$ given in publications from year T:

$$R(T, t) = N(T)\, N(t)\, P(T, t)\,, \tag{4}$$

where $N(T)$ is the number of publications in year T, i.e., the number of citing publications; $N(t)$ is the number of publications in year t ($t \leq T$, i.e., all years that can be cited, and thus the years from which the references may originate); $P(T, t)$ is the probability that a publication of year T will cite a publication of year t, which then becomes a reference with age $\tau = T - t$.

We now consider two cases:

I. For the period $T_0 \leq t < T_1$:

This is the 'final increase' period. In the line of our above formulated hypothesis, we state that the probability function $P(T, t)$ can be expressed by:

$$P(T, t) = A\, e^{-b(\tau - \theta)} \tag{5}$$

with, as given above, $\tau = T - t$, and $\theta = T - T_1$; furthermore $A < 1$.

The factor $\tau-\theta$ instead of τ is necessary, as the exponential ageing of references does not start in year T, but in year T_1, the 'peak-year'.

II. For the period $t < T_0$:

This is the period 'before the final increase'. Again in line with our hypothesis we state that in this case the probability function $P(T, t)$ will be in first approximation:

$$P(T, t) = B\,, \tag{6}$$

which simply means a constant value, with B<A.

The ratio of the number of references $R(T, t)$ in the period $T_0 \leq t < T_1$ (case I) to the number of references extrapolated from the period $t < T_0$ into the period $T_0 \leq t < T_1$ gives us an expression for the relative size of the 'final increase'. With help of Eqs (4–6) we find

$$r(T,t) = \frac{A\,N(T)\,N(t)\,e^{-b(\tau-\theta)}}{B\,N(T)\,N(t)} = (A\,/\,B)\,e^{-b(\tau-\theta)}\ . \tag{7}$$

We can evaluate Eq. (7) in the two limiting situations.

First at the position of the peak: $t = T_1$ thus $\tau-\theta = (T-T_1)-(T-T_1) = 0$, which gives

$$r(T, t) = (A/B) > 1, \text{ as } B < A\,. \tag{8}$$

Second, the position where the 'final increase' starts: $t = T_0$ thus $\tau-\theta = (T-T_0)-(T-T_1) = T_1-T_0$, which gives

$$r(T, t) = (A/B)\ \exp\{-b(T_1-T_0)\}\ . \tag{9}$$

As at this position $r(T, t) = 1$ by definition, we find

$$A\exp\{-b(T_1-T_0)\} = B\,. \tag{10}$$

The model thus satisfies the requirements we demand, namely an explanation and description of the 'final increase' in the period $T_0 \leq t < T_1$. But it also reveals that for the period $t < T_0$ the very largest part of the entire period – we may consider, in good approximation, the measured age-distribution as a direct fingerprint of the growth of science, according to

$$R(T, t) = B\,N(T)\,N(t)\,,$$

as B is a constant with a value < 1, and N(T) is a fixed number for a given publication year T (here: 1998), which indeed is equivalent to Eq. (2).

We already indicated that the above model has to be considered as a first but good approximation. From the empirical data (Fig. 1) the growth rate of science can be found to be 10% per year. Further refinements, particularly at a lower aggregation level, i.e., growth phenomena in specific fields of science, should take the possibility of a direct influence of growth on ageing as a 'higher order' effect into account. An important discussion of such an effect is given by *Egghe* et al.[14,15]

Mathematical model for fractal differentiation

From the above it follows that in good approximation the growth of science can be described for not too long periods of time (say, 50 years) with a simple exponential function. Using Eqs 1, 2, and 3h and taking an arbitrary starting point at t_0, we find for the total number of publications (which we consider as a realistic representation of the 'size of science') for $t > t_0$:

$$N(t) \cong N(t_0) \exp a\Delta t, \text{ with } \Delta t = t - t_0 \ . \tag{11a}$$

Now we develop our model of growth and differentiation by making the following three steps.

1. In any complicated, large system with very many elements, total mutual connectivity of all elements is impossible, in other words: *any real system is not totally connected*. This implies that 'local interactions' normally dominate the system.

Therefore we assume (Assumption 1) that the growing science system is built up of many sub-systems i (think of major disciplines, fields of science, etc.) each having their own 'size' $N_i(t)$ with an exponential growth process similar as given by Eq. (11a):

$$N_i(t) \cong N_i(t_0) \exp a_i\Delta t \ . \tag{11b}$$

A striking example of this exponential growth of 'scientific sub-systems', i.e., major disciplines, is given by the increase in the last thirty years of the number of publications in physics and related fields measured on the basis of the large international database INSPEC/Physics Abstracts.[7]

We have to prove that under the above condition, the system as a whole exhibits, at least in good approximation, an exponential growth too. Thus, for science as a whole we rewrite Eq. (11a) as

$$\begin{aligned} N(t) &= \Sigma_i N_i(t) = N_1(t) + N_2(t) + N_3(t) + \ldots = \\ &N_1(t_0)\exp(a_1\Delta t) + N_2(t_0)\exp(a_2\Delta t) + N_3(t_0)\exp(a_3\Delta t) + \ldots = \end{aligned} \tag{12}$$

As we are free to choose the order of this summation, we take the sub-system with the largest growth-coefficient as the first term $i = 1$, with $a_1 > a_i$ $\{i \neq 1\}$, which allows us to write Eq. (12) in the form

$$N(t) = N_1(t_0)\exp(a_1\Delta t).[1 + \{N_2(t_0)/N_1(t_0)\}\exp((a_2-a_1)\,\Delta t) + \ldots\ldots]. \quad (13)$$

For large t ($t \rightarrow \infty$) this expression will approach to

$$N(t) = N_1(t_0)\exp(a_1\Delta t)\,, \quad (14)$$

which proves that the system as a whole will indeed behave in an exponential way, however increasingly dominated by the fastest growing sub-system, as is to be expected.

2. So far we used 'ageing' in a rather broad, literature-related sense: the diminishing preference of readers for older literature, and given our hypothesis only relevant for the recent period from T_1 to T_0. In other words, we see ageing in this sense just as a temporally restricted and clearly located 'disturbance' which can be accounted for by a rather simple mathematical treatment of the empirical data. This model then allows us to interpret the largest part of these empirical data as – in good approximation – the result of growth.

Now we look at a more specific type of ageing which is not – like the above literature ageing – a disturbance in the empirical registration of growth, but a more genuine, 'biological' type of ageing, and directly related to growth phenomena. We here mean the age of important breakthroughs, discoveries that are almost always the starting points of new fields of research, of new 'sub-systems' as we have called them earlier in this paper. For instance, ten years ago discovery A was made. Thus, discovery A has by definition the age of ten years and the research field (sub-system of science) resulting from this discovery has a lifetime of ten years. As discussed earlier, we assume that these new sub-systems will grow in first approximation, at least during a substantial part of the time after the start, in a 'biological', i.e., exponential way.

Now we assume (Assumption 2) that the growing sub-systems are the cradles of further, new developments, and that the *probability* of new breakthroughs is – again in first approximation – *proportional to the size* of the sub-system. On their turn, the new breakthroughs give rise to further new field of research, and so on. We developed the following mathematical model to describe the dynamics of this process.

Between t_0 and the time 'now', t, we consider time t_1. By taking $t-t_1 = x$ we introduce the age x of a new development originating at t_1. If we conceive fields of science as coherent sets of publications developing in time like living organisms and originating in one of the many exponentially growing sub-systems, we can use an 'ecological' model (*Pielou*[16]) of science in calculating the age-distribution, which goes as follows. We take the first sub-system as an example.

At any time t_1 there is a probability of a new development, a 'birth' or 'creation' of a new field, proportional (factor $C < 1$) to the (time-dependent) size of the 'mother-field' $\{i\}$ in which the new development takes place. We therefore may write the number of new developments originating from the 'mother-field' at time $t_1 (= t-x)$ as

$$C\,\mathrm{N}_i(t-x)\,.$$

From Eq. (11b) follows

$$\mathrm{N}_i(t) = \mathrm{N}_i(t_1)\exp\{a_1(t-t_1)\} = \mathrm{N}_i(t-x)\exp(a_i x)\,, \tag{15}$$

which yields

$$\mathrm{N}_i(t-x) = \mathrm{N}_i(t)\exp(-a_i x)\,, \tag{16}$$

so that the probability $\mathrm{c}(x)\mathrm{d}x$ with which the mother-field generates new developments in the age-interval $\{x, x+\mathrm{d}x\}$ at time t is given by

$$\mathrm{c}(x)\,\mathrm{d}x = \{C\,\mathrm{N}_i(t-x)/\mathrm{N}_i(t)\}\,\mathrm{d}x = C\exp(-a_i x)\,\mathrm{d}x\,. \tag{17}$$

Assuming that the probability that at least once a new development will occur, we evaluate the integral of Eq. (17) with 0 and ∞ as limits of integration:

$$\int_0^\infty C\exp(-a_i x)\,\mathrm{d}x = 1\,. \tag{18}$$

Evaluating this integral, we easily find that $C = a_i$.

We see from the above that an exponential growth of fields of science yields an exponential age-distribution of new developments in science, given by Eq. (17).

3. The last step brings us to differentiation, more precisely, the size-distribution of all sub-systems within the system as a whole. The intriguing aspect of our approach is that a sub-system size-distribution can be derived from the two above discussed, time-dependent phenomena growth and ageing. Following our assumption that new fields, originating from a mother-field, grow 'biologically', i.e., in first approximation with a simple exponential function, we can write the 'size' (operationalised, as everywhere in this model, in terms of number of publications) for these newly created fields again in a

similar way as in Eq. (11b). In line with the discussion in the foregoing step, we write the size of a new field in terms of its age:

$$N_n(x) = \exp(a_n x) \tag{19}$$

(starting with one publication, i.e., $N_n(0) = 1$, which is only for the ease of calculation, $N_n(0)$ may be other positive values).

Now we have an *age probability distribution* function for new developments at time t (given by Eq. (17) *and* Eq. (19) for the *size-distribution* as a function of age. Using similar considerations as *Fermi*[17] and *Naranan*[18] (and even much older literature given by *Naranan*)[19] we can now calculate a *probability distribution function for the size* of new developments at time t.

By writing $N_n(x) = N$ for simplicity we have from Eq. (19):

$$x = (1/a_n)\, ln\, N\,, \tag{20}$$

which immediately yields

$$dx/dn = 1/a_n N\,. \tag{21}$$

We write the age probability distribution function Eq. (17) as

$$dp(x) = c(x)dx, \text{ hence } dp(x)/dx = c(x),$$

and make the transformation to the size probability distribution by

$$dp(N)/dN = \{dp(x)/dx\}\{dx/dN\}\,.$$

Substituting Eqs (20) and (21) in Eq. (17) and taking $a = a_1$ and $b = a_n$ for simplicity, we find for the number of new developments with size in the interval {N, N+dN}:

$$dp(N)/dN = (1/bN)\, a \exp\{-a(1/b)\, ln\, N\} = (a/b)\, N^{-(1+a/b)}\,, \tag{22}$$

which clearly is a *power-law* size-distribution.

Now we have made with the above model the link to our earlier empirical findings,[20] where we showed on the basis of co-citation clustering analysis that science (more precisely: the total ensemble of scientific literature) is 'fragmented' according to a power-law distribution. In line with this earlier work, Eq. (22) can be rewritten in terms of a *size-rank distribution* where $n(r)$ is the size with rank r given by

$$r = \int_n^\infty \{dp(N)/dN\}\, dN = N^{-a/b}\,,$$

hence

$$N(r) \sim r^{-\gamma}, \text{ with } \gamma = b/a.$$

We identified in the earlier work the power-law exponent γ as the reciprocal value of the fractal dimension of the fragmentation or differentiation of the 'scientific ecosystem', i.e., $\gamma = 1/D$.

The importance of the model presented in this paper is that it explains the fractal dimension, which is a momentary, 'static' description of the structure of science (again: as reflected by the scientific literature), in terms of two time-dependent processes with form the basis of our model: ageing and growing. This is clearly visible in the power-law exponent $\gamma = b/a$, since $1/b$ can be considered, given Eq. (19), as a *characteristic growth-time* of a new development, and $1/a$, given Eq. (17), as a *characteristic life-time* of a new development.

We know from the co-citation work[20] that the finer we observe new developments (i.e., at increasingly lower levels of aggregation), the lower the power-law exponent γ and thus the higher the fractal dimension D becomes. This observation is – at least qualitatively – in agreement with the model presented in this paper. Going from a high level of aggregation – where we observe only large, 'established' and thus relatively old developments – toward finer structures we observe more and more 'younger' developments, which implies that the characteristic life-time $1/a$ becomes increasingly smaller (for instance: 3 years) as compared to the characteristic growth-time $1/b$ (for instance: size-doubling after 7 years), which means a decreasing γ.

Conclusions

On the basis of the measured time-dependent distribution function of references in recent (1998) scientific publications, over a total period of two centuries, we have made a number of observations leading to the mathematical formulation of ageing and growth of science. This model has some important novel features.

First, the method to 'filter out' the ageing process is a new approach to find a better description of the growth of science. Second, the concept of a dynamic system consisting of growing sub-systems forms the basis of an approach to the differentiation of science into the very many different fields, sub-fields, themes. Growing sub-systems drive as it were an exponentially age-distributed source of new, mostly relatively small developments. On their turn, these new developments start to grow. The combination of both time-dependent processes reveals a power-law distribution of the 'size' of the sub-systems that essentially form the structure of science. Thus, we derive from our empirical findings a fractal differentiation model of science, which is in agreement with earlier empirical findings[20] concerning the clustering of scientific literature.

Power-law phenomena are typical in bibliometric analysis.[19] To what extent we can deduce from our model more, most, or may be all of these phenomena, remains to be seen and is object of our current research. The same has to be said for the consequences of our model in relation to self-organisation of large systems. More and strikingly universal power law phenomena in the growth dynamics of complex systems are recently reported by the group of *Stanley*.[21,22,23] Fractal structure and the associated power-law distribution functions are known to be closely related to the problem of self-organisation, particularly the propagation of order in large systems (for the dramatic consequences in biological systems, see *Kauffman*).[24] As we already discussed, in any real system there will be no total connectivity of all the parts and elements of the system, as this situation would over-complicate the system. Local interactions will normally dominate the system, therefore the system can be characterised as an extended web-structure, and this is in line with our empirically supported model of clustered sub-systems.

We think that our observations reinforce the idea of science as a largely self-organising 'cognitive ecosystem'. The fractal distribution provides a measure of the diversity of this cognitive ecosystem, that is, the distribution of scientists among larger and smaller scientific 'species' (disciplines, fields, sub-fields, themes, etc.).

Often self-organisation is explained with dynamic properties of a complex system characterised by *self-organised criticality*,[25,26] in which 'events' occur with power-law statistics in time and space. But in this paper we proved that it is not necessary to start a priori with the assumption of self-organised criticality to find fractal distributions. Or can fractal distributions always be seen as a fingerprint of systems in a critical state, i.e., self-organised order on the verge of chaotic behaviour?

For the science system this is a most interesting question. In critical systems events that may differ in many magnitudes of size (e.g., earthquakes, avalanches, mass extinction of species, and now perhaps scientific breakthroughs) occur on the basis of the same mechanisms (i.e., there is no special, 'additional' cause necessary for a major earthquake as compared to a small one), and the probability of these events is only dictated by a power-law. Such considerations would have major implications for our ideas on the 'makebility' of scientific breakthroughs and innovations. But these findings also have remarkable philosophical implications: smaller and larger discoveries are the 'rule' in our self-organising cognitive eco-system, and it is impossible to regard the larger discoveries as essentially different, in terms of the system dynamics, from the smaller breakthroughs. But then the ground under the paradigm theory of Kuhn disappears. There is no 'normal' science alternated with well-defined periods of 'revolutionary' science in which new paradigms start to dictate the rules. Science is *always* revolutionary, but by the typical statistics of complex systems, there are mostly smaller and only rarely big breakthroughs.

*

The author gratefully acknowledges the Netherlands Organisation for Scientific Research (NWO) for financing substantial parts of the building and further development of our bibliometric data-system. Particularly the work of Peter *Negenborn* of our institute is of great value for the advancement of this data-system. I thank him for his efforts to create sets of the reference age-distributions that form the basis of this study.

I thank Professor Johan *Grasman* of Wageningen University for stimulating discussions on several important mathematical topics concerning growth phenomena. I am also grateful to the two referees for their critical remarks and most valuable suggestions.

References

1. PRICE, D. J. DE SOLLA, *Little Science, Big Science*. New York: Columbia University Press, 1963.
2. PRICE, D. J. DE SOLLA, Networks of scientific papers, *Science*, 149 (1965) 510–515.
3. PRICE, D. J. DE SOLLA, *Little Science, Big Science, ...and Beyond*. New York: Columbia University Press, 1986. (Revised and enlarged edition of Ref. 1 including also Ref. 2).
4. GRIFFITH, B. C., Derek Price's puzzles: Numerical metaphors for the operation of science, *Society of the Social Studies of Science (4S) Price Memorial Lecture*, 1987.
5. MACRAE JR., D., Growth and decay curves in scientific citations, *American Sociological Review*, 34 (1969) 631–635.
6. GRIFFITH, B. C., P. N. SERVI, A. L. ANKER, M. C. DROTT, The aging of scientific literature: a citation analysis, *Journal of Documentation*, 35 (1979) 179–196.
7. VAN RAAN, A. F. J., Further observations on growth and ageing of science, to be published.
8. MOED, H. F., R. E. DE BRUIN, TH. N. VAN LEEUWEN, New bibliometric tools for the assessment of national research performance: database description, overview of indicators and first applications, *Scientometrics*, 33 (1995) 381–422.
9. VAN RAAN, A. F. J., Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises, *Scientometrics*, 36 (1996) 397–420.
10. LAIRD, A. K., Dynamics of Tumor Growth, *Brit. J. Cancer* 18 (1964) 490–502.
11. GOMPERTZ, B., On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies, *Phil. Trans. Roy. Soc.*, 115 (1825) 513–585.
12. EGGHE, L., I. K. RAVICHANDRA RAO, Classification of growth models based on growth rates and its applications, *Scientometrics*, 25 (1992) 5–46.
13. ROUSSEAU, R., Double exponential models for first-citation processes, *Scientometrics*, 30 (1994) 213–227.
14. EGGHE, L., On the influence of growth on obsolescence, *Scientometrics*, 27 (1993) 195–214.
15. EGGHE, L., I. K. RAVICHANDRA RAO, R. ROUSSEAU, On the influence of production on utilization functions: obsolescence or increased use? *Scientometrics*, 34 (1995) 285–315.
16. PIELOU, E. C., An Introduction To Mathematical Ecology. New York: John Wiley & Sons, 1969.
17. FERMI, E., On the origin of the cosmic radiation, *Physical Review* 75 (1949) 1169–1174.
18. NARANAN, S., Bradford's law of bibliography of science: an interpretation, *Nature*, 227 (1970) 631–632.
19. NARANAN, S., Power law relations in science bibliography: a self-consistent interpretation, *Journal of Documentation*, 27 (1971) 83–97.
20. VAN RAAN, A. F. J., Fractal dimension of co-citations, *Nature*, 347 (1990) 626.

21. LEE, Y., L. A. N. AMARAL, D. CANNING, M. MEYER, H. E. STANLEY, Universal features in the growth dynamics of complex organisations, *Phys. Rev. Lett.*, 81 (1998) 3275–3278.
22. AMARAL, L. A. N., S. V. BULDYREV, S. HAVLIN, M. A. SALINGER, H. E. STANLEY, Power law scaling for a system of interacting units with complex internal structure, *Phys. Rev. Lett.*, 80 (1998) 1385–1388.
23. PLEROU, V., L. A. N. AMARAL, P. GOPIKRISHNAN, M. MEYER, H. E. STANLEY, Similarities between the growth dynamics of university research and of competitive economic activities, *Nature*, 400 (1999) 433–437.
24. KAUFFMAN, S. A., *The Origin of Order. Self-Organization and Selection in Evolution*. New York/Oxford: Oxford University Press, 1993.
25. BAK, P., C. TANG, K. WIESENFELD, Self-organized criticality, *Physical Review*, A, 38 (1988) 364–374.
26. BAK, P., K. CHEN, Self-organized criticality, *Scientific American*, January 1991, 26–33.

(Received October 27, 1999)

Address for correspondence:
ANTHONY F. J. VAN RAAN
Centre for Science and Technology Studies, University of Leiden,
Wassenaarseweg 52, P.O. Box 9555, 2300 RB Leiden (The Netherlands)
E-mail: vanraan@cwts.leidenuniv.nl