J. Eng. Technol. Manage. 32 (2014) 185–205

Contents lists available at ScienceDirect

Journal of Engineering and Technology Management

journal homepage: www.elsevier.com/locate/jengtecman

# Tipping points in science: A catastrophe model of scientific change

S.W. Cunningham *, J.H. Kwakkel

*Faculty of Technology, Policy and Management, Delft University of Technology, 2600 GA Delft, The Netherlands*

ARTICLE INFO

*Article history:*
Received 4 February 2013
Received in revised form 22 January 2014
Accepted 23 January 2014

*Keywords:*
Scientometrics
Diffusion modeling
Scientific growth
Logistic
Catastrophe

ABSTRACT

In this paper we discuss the capabilities for scientific knowledge to demonstrate explosive growth in short periods of time. In one notable example the field of engineering and technology management grew more rapidly in the 4 years after 1980 than it was expected to grow for the next 40 years. We provide 22 examples drawn widely from science, demonstrating that this phenomena is pervasive throughout science. We propose a new model, based on the idea of folds from mathematical catastrophe theory, a phenomenon that is more popularly known as tipping points. This model is then fit using non-linear regression in the presence of Poisson noise. While the tipping point does not occur in all fields of science, in those cases where it does occur the resultant model overwhelmingly supports the idea of catastrophic growth within scientific knowledge. We describe the differential equations underlying the fold catastrophe and relate these equations to a process of communication and interaction. We relate this dynamic to other word of mouth models such as the Bass diffusion model. We further discuss why scientific, and to a lesser extent news, articles are subject to this behavior while the same phenomenon is unlikely to occur when solely measuring the sales of a physical product. We provide evidence of the phenomenon in one brief sociological sketch of scientific activity. Finally, we discuss the relevance of the model in terms of innovation forecasting. In particular, we evaluate the possibility for ex ante anticipation of the bifurcation point.



* Corresponding author. Tel.: +31 015 27 87187; fax: +31 15 27 87187.
*E-mail addresses:* s.cunningham@tudelft.nl (S.W. Cunningham), j.h.kwakkel@tudelft.nl (J.H. Kwakkel).

0923-4748/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jengtecman.2014.01.002

## Introduction

The field of scientometrics attempts to create both static and dynamic measures of knowledge production. Previous scientometric research into dynamic measures of knowledge production has examined a range of alternative measures including word usage (Noyons and van Raan, 1988), citation usage (Boyack et al., 2007; Garfield, 2004), and patterns of collaboration (Reid, 1997). Scientometricians often use a sociology of science perspective when modeling the growth of knowledge (Zitt, 1991). This perspective, in turn, is strongly influenced by semiotics and linguistics (Rip and Courtial, 1984; Zitt, 1991). Our principal concern in this paper is the quantitative modeling of the growth of scientific knowledge. This concern in scientometrics dates to the onset of the field, and is perhaps best credited to the great philosopher and sociologist of science Derek de Solla Price. In this paper we discuss the capabilities for scientific knowledge to demonstrate explosive growth in short periods of time.

Studying the dynamics of science can be done at different levels of granularity. For example, one can look at the past development of a scientific field, at the research front currently faced in a particular research field, alternatively, one can study the broader disciplinary structure of the sciences, or instead in much more detail delve into a subfield and its dynamics. That is, one can study, the disciplinary structure, the fields within a discipline, the subfields that constitute a field, and research topics in a particular subfield (van den Besselaar and Heimeriks, 2006). Much of the current literature on science dynamics is of the macro-dynamic character. This literature is significant both for policy as well as of science. de Solla Price's (1961, 1963). contributions are of this character, and the work of Katz and Hicks (Hicks and Katz, 1996) representative of the policy relevant strand of research.

De Solla Price, musing upon a complete collection of the *Philosophical Transactions of the Royal Society of London*, observed that science has been growing exponentially. These observations were first noted in a 1963 book and then more famously in his 1963 work *Little Science, Big Science* (de Solla Price, 1961, 1963). This later book, which endorsed quantitative methods for analyzing science, fortuitously occurred at the birth of a new discipline of scientometrics. Not surprisingly, given its time and its content, the book became a citation classic for the field (de Solla Price, 1983).

De Solla Price, who was trained as a physicist, well knew the consequences of unending exponential growth. In de Solla Price's own words "the exponential growth business needled me a lot (de Solla Price, 1983)." He therefore postulated that eventually science must reach a steady state. His argument is that there is a finite population of potential scientists. Further, there will be decreasing returns to productivity as increasing numbers of scientists are trained and recruited by society. This occurs because of a gradation in scientific talent. The first scientists are the most talented, and are therefore trained in the most cost-affordable manner, producing the greatest marginal gains in scientific output. Recruitment and training for subsequent scientists becomes more difficult, with lesser gains in output achieved a greater expense. Publications here are taken as a measure of scientific output. De Solla Price, like many others after, understood the limitations of publication as the sole measure of scientific output. Nonetheless he convincingly describes the use of publication output as a partial indicator of scientific progress.

The logistic curve is an obvious choice for modeling diffusion limited growth. The dynamics of the logistic are compounded from two processes. The first process involves growth in proportion to an existing population. This is variously known as the dispensatory or replication process. The second process modifies the growth so that, as the population approaches its limit or stable carrying capacity, the growth rate approaches zero. The second process is known as the compensatory or inhibiting process (Miranda and Lima, 2010). The combination of these two processes results in the characteristic S-shape curve. Growth starts low, expands rapidly, passes through an inflection point, slows down and asymptotically approaches a saturation limit. See Fig. 1 for an example of how this might apply to the dynamics of publication output.

Verhulst first studies this dynamic in light of population biology (Verhulst, 1838). These dynamics were later rediscovered by Pearl (Pearl and Reed, 1920). Fisher and Pry described technological substitution behavior in light of logistic growth, although without direct reference to the earlier ecological applications (Fisher and Pry, 1971). Pearl and Fisher Pry curves are one form of trend extrapolation among many referenced by a premier text on technological forecasting.

Particularly noteworthy for our purposes are those papers which attempt to forecast growth in science and technology (Bengisu and Nekhili, 2006; Daim et al., 2006). This work serves a useful

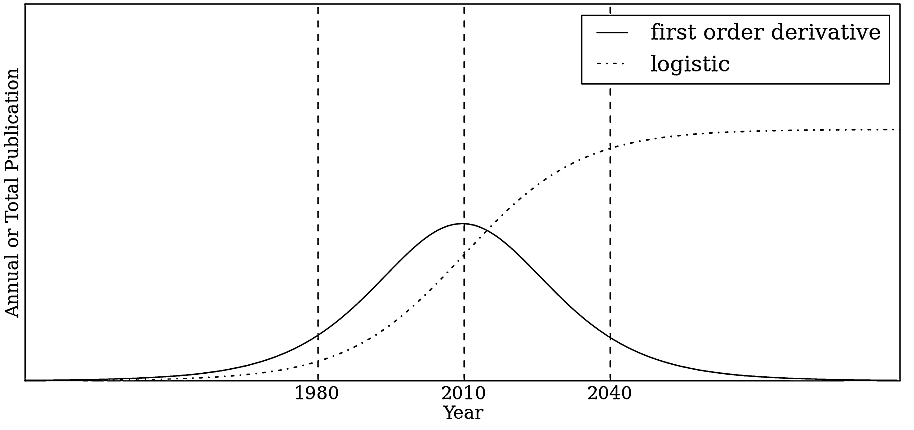


**Fig. 1.** Logistic curves and publication outputs.

function for identifying new and emerging areas of research well prior to market commercialization. A principal inspiration for these researchers are existing models of technology diffusion. The various models used in these different field of research often postulate that growth is either linear, exponential or S-shaped in character (Bass, 1969; Porter et al., 1991; Roper et al., 2011). These models thus postulate a continuous growth, which, in case of exponential models constantly accelerates, and in case of logistic models, slowly decelerates again.

These basic models thus cannot account for the discontinuous and sudden growth that is observed in the real world. In one notable example, the field of engineering and technology management grew more rapidly in the 4 years after 1980 than it was expected to grow for the next 40 years (Cunningham and Kwakkel, 2011). Other examples can easily be found across the sciences. Various explanations for this discontinuous behavior can be offered. Adopting a Kuhnian perspective on the dynamics of science, terminology merely reflects the wider dynamics. Thus, scientific revolutions are accompanied by sudden increases in very specific terminology, associated with the new paradigm. Another, more bottom up explanation is that scientists are searching for words and phrases to describe their work. In particular in a new area of study, a variety of words and phrases can be used to denote roughly the same idea or object. Scientist might start to discuss harmonizing terminology, to facilitate the exchange of ideas, leading to small shifts in which terms are being used. At some point, one of the candidate words or phrases reaches a critical mass, sufficient to swamp the competition. Resulting in the rapid adoption of that word or phrase. Yet another explanation is the fact that new terminology might bring together different research fields that before that point worked in isolation using their own terminology. The new terminology is then able to tap into these different research fields, facilitating the exchange of ideas across these fields and leading to a quick spread over these fields that up until then used their own research field specific terminology. A fourth explanation for this dynamic comes from the distortive effects of national science funding. Funding agencies put out calls inviting research proposals on particular topics. Scientists will write their proposals to fit these calls, adopting the terminology of the call to describe their own research. A prime example of this dynamic comes in the form of the various "nano" related terms. There is ample evidence that a large part of nanotech funding goes to well established research areas that had their own terminology, which because of funding reasons is being rebranded as nanotech research.

Alternative models may be equally effective summarizations of empirical data. For instance in the early stages of growth, hyperbolic, exponential, and logistic growth may be indistinguishable despite having very different growth dynamics, and very different results even over the medium term (Fig. 2). Theories of underlying growth, as well as empirical evidence in support of the theory, are necessary to choose between alternative explanations of phenomena. In the following paragraphs we consider three epiphenomena of scientific growth. When these phenomena are present, logistic models are poor representations.

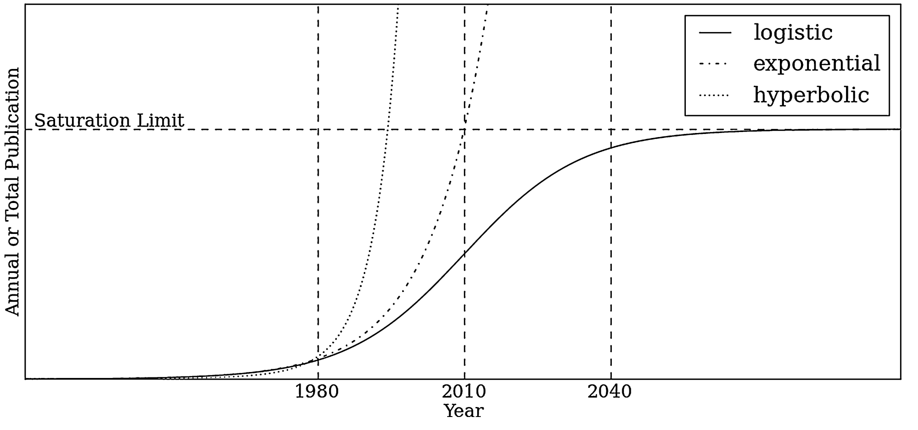


**Fig. 2.** Coincidence of alternative publication models.

The first epiphenomena in the empirics of scientific knowledge is coalition growth. In the world science example given above, there was a period of very rapid scientific expansion starting in the 1940s. Scientific output grew by an order of magnitude in 25 years. Previously such growth would have taken 75 years (see Fig. 2). It is telling that faster than exponential growth is known as hyperbolic or "coalition growth" and such growth has been characteristic of human population growth for a long span of time (von Foerster and Mora, 1960). Hyperbolic growth is enabled by positive returns to scale from cooperation. This phenomenon has previously been noted in science, where a particular manifestation is known as the Matthew effect (Merton, 1960). Others have noted how urbanization provides a dramatic multiplier on economic activities, presumably facilitated by intense interaction in a circumscribed space (Bettencourt et al., 2007). Such growth is inexplicable given the dynamics of the logistic. In logistic growth there is always a deceleration, but never an acceleration in absolute numbers over time.

The second epiphenomena is two-level growth. The limit or carrying capacity of science is not a fixed quantity. De Sola Price fixed scientific carrying capacity at a fixed proportion of population, acknowledging that it may be many years still before full recruitment of scientists is realized, and therefore a steady state achieved. Further, although population growth has slowed it has not reached a steady state. The production of Ph.D.s is to a large degree, a phenomena of the developed world, suggesting that Ph.D. researchers are intensive on public goods such as national science funding and a university system. Thus the economy could also form a basis for fixing the carrying capacity of science. However like world population the world economy is not at a steady state. Both population and the economy are time-varying quantities. Thus the assumption of a finite pool of knowledge given a fixed resource base is inappropriate. However this assumption of a fixed carrying capacity is inherent in the logistic growth model.

The third epiphenomena are intellectual migration. Researchers are broadly capable, and thus can select from a range of different research specialties as need, circumstance, or opportunity require. This is very different from the assumptions of the logistic equation where a given species has a single resource base. When the resource base is exhausted, growth is permanently curtailed. The assumption of a single species with a single resource base is relaxed when modeling ecological webs. The resulting models, like actual species distributions over time and place, may display punctuated equilibria (Hubbell, 2001). A further manifestation of intellectual migration is achieving a critical mass of scientists to solve a difficult problem. We address this point further below.

For three reasons then, the logistic model should be questioned. Scientific growth involves coalition behavior, where there is positive returns to scale. Scientific growth is in itself dependent on other quantities of population and economy. Scientific growth is conditioned on intellectual migration, and epistemic relationships between fields of study. Empirical findings may be effectively and naively summarized in any number of ways. A theoretical account of the underlying dynamics of the trend is the only principled means of choosing between alternative explanations.

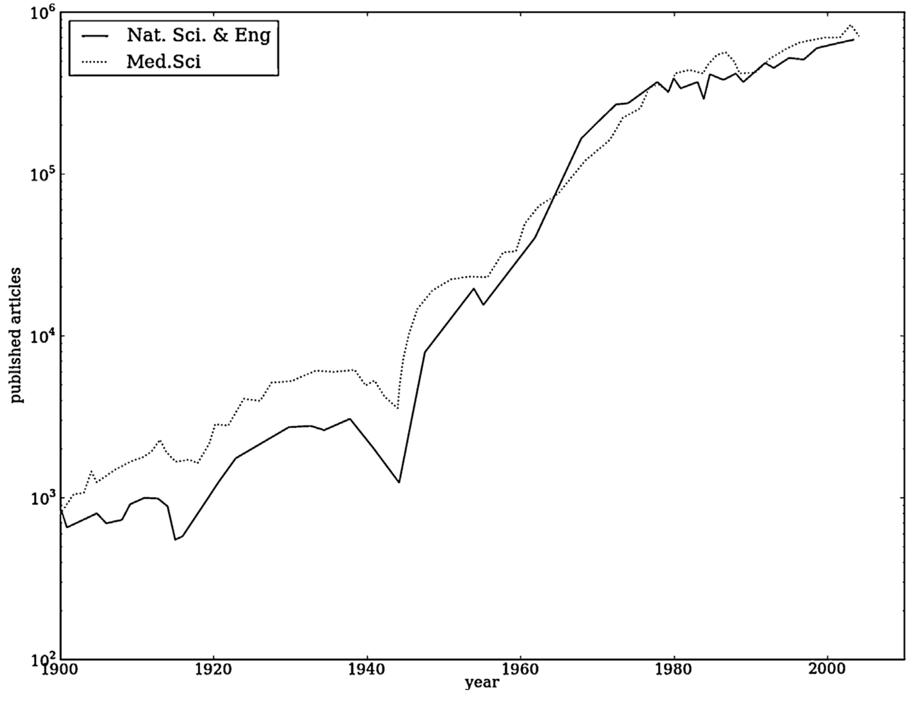


**Fig. 3.** Annual number of articles for medical fields and natural sciences and engineering.

Fig. 3 shows a very typical dynamical trajectory of annual number of articles in two major fields of science and technology. The original data is reported in Lariviere et al (2008) and further discussed in Han et al. (2010). Note the only superficial relationship to a logistic, or other S-shaped curve. The chief departure of the curve is either the very slow growth from initial conditions, or the extremely rapid growth which occurred between 1950 and 1970. Either (but not both) phenomena are consistent with logistic growth. De Solla Price observes qualitatively very similar trends in publication, reporting total rather than annual output (de Solla Price, 1963). In this paper we examine scientific growth at a much narrower level, corresponding to the usage of single words or phrases by scientific authors. The same qualitative dynamics are apparent, although the inflationary period displays much more extreme behavior. We identify the field/subfields through co-word analysis. The underlying rationale of this approach is that co-word analysis gives direct access to the research topics in terms of concepts as used by the researchers (van den Besselaar and Heimeriks, 2006).

A final point of confusion with the logistic modeling of scientific growth is whether the model is appropriate at the level of single papers, or cumulative numbers of papers published. Both styles of modeling are present in the literature, yet both cannot be true as they present mutually contradictory hypotheses of growth. A model of total publication is intended in de Solla Price's original vision (de Solla Price, 1963). Logistic growth in total papers implies absolute long-term declines in year over year production of scientific literature, something which is comparatively rare in the data. Other authors use the logistic as a model of annual not total publication (Bengisu and Nekhili, 2006; Daim et al., 2006). The two models are mutually inconsistent; both cannot be true unless they are describing two distinct publication phenomena.

We propose a new model, based on the idea of folds from mathematical catastrophe theory. This model postulates a slow growth process leading up to a point at which the growth curve suddenly changes dramatically. This point at which the growth shows a sudden discontinuous jump is more popularly known as a tipping points. This insight can be used for various purposes ranging including individual scientists updating their vocabulary, funding agencies tailoring research calls, and technology management. In section 'Method', we present the model. In Section 'Application of the model', we select a set of cases and apply the model to these cases. Section 'Discussion' contains a discussion of the results. Such a model can be used in the context of monitoring scientific developments, and can provide early warning that a particular word or phrase has passed a tipping point. Section 'Conclusion' presents the conclusions, and its relevance for the broader field.

## Method

### *The origins of catastrophe*

Catastrophe theory in mathematics is a branch of bifurcation theory, which is part of the wider domain of dynamical systems. Bifurcation theory studies how massive changes can arise out of small shifts in conditions. Catastrophe theory studies the situation where the long-run stable equilibrium is identifiable. The point at which the system suddenly shifts to a different type of behavior is also, more popularly, known as a tipping point. Different types of catastrophes have been identified, based on the number of control parameters in the equation. For example, if there is only a single control parameter, one speaks of a fold catastrophe, while if there are two control parameters, one speaks of a cusp catastrophe.

Tipping points have become popular, among others, through the work of Malcom Gladwell (Gladwell, 2002). There are also studies in the complexity sciences concerning tipping points in the context of complex networks. Tipping point behavior has for example been found in scale free networks. Scale free networks are networks that have no characteristic scale. They are characterized by power law distributions. That is, the degree of the nodes in the network follows a power law. The result of this is that scale free networks have a few nodes with a very high degree, the variance of the degree of the nodes is very large, the network is self-similar, and has the small world property.

Tipping point behavior in this context, are points at which some process starts to either increase dramatically, or can completely disappear. Both dynamics are driven by a positive feedback loops. A prime example of such behavior is the accumulation of citations (Mitchell, 2009). Tipping points are easily introduced into ecological models by means of the Allee effect. The Allee effect a positive relationship between density and population growth. Such a relationship may be related to herding, the benefits of a diverse gene pool, or the mutualism of highly communal or social species (Mendez et al., 2011).

Our approach examines local dynamics in the context of gradual systemic change. Slow-fast vector fields are one approach for understanding dynamical change in "systems of systems." A slow-fast vector field has the form

$$\varepsilon\dot{x} = f(x, y, \varepsilon) \tag{1}$$

$$\dot{y} = g(x, y, \varepsilon) \tag{2}$$

with $x \in R^m$ and $y \in R^n$. The variable $x$ is the so-called fast variable, and the variable $y$ is the slow variable. One such dynamical system in $x$ and $y$, which gives rise to prominent fold catastrophes, is represented by the following model

$$\dot{x} = y - x^2 + \xi \tag{3}$$

$$\dot{y} = az + bx + \eta \tag{4}$$

$$\dot{z} = 1 + \zeta \tag{5}$$

We wish to understand the dynamic behavior of $x$, $y$ and $z$ in the presence of $\xi$, $\eta$, $\xi$,, or higher-order terms. This system can be written to first approximation with the following set of equations

$$\begin{aligned} \dot{x} &= y - x^2 \\ \dot{y} &= az - x \\ \dot{z} &= 1 \end{aligned} \tag{6}$$

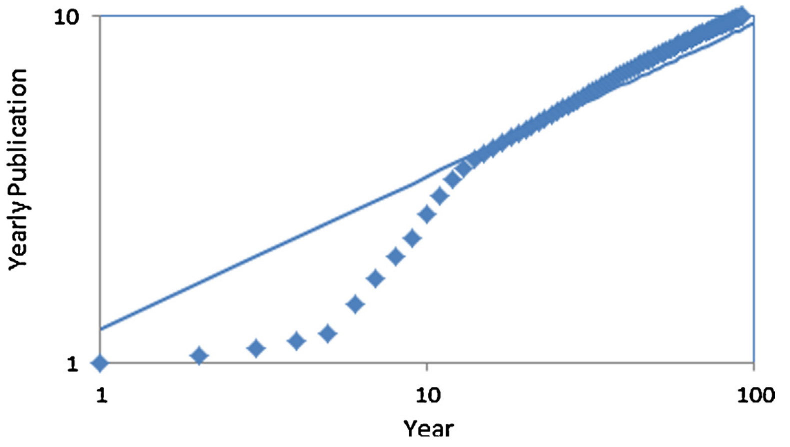


**Fig. 4.** Fold dynamics in scientific production.

Dynamical preserving transformations of these systems are studied in Arnold et al. (1994) and Guckenheimer and Haiduc (2003). Arnold is well-known for earlier contributions in the field of catastrophe theory (Arnold, 1984). A sample dynamic trajectory of this system of equation is given in Fig. 4.

This is a solution to the differential equations given in (1), when $a = 1$. Research production starts low, but in a few years dramatically ramps upwards, ultimately resulting in a high and constant rate of growth. Note that the growth rate in this figure is proportional to the square root of time. We will return to this in a simple time series model of publication behavior, below. In an actual time series the date of bifurcation might vary, and the initial and tipping point growth rates might also vary. Further, the bifurcation need not be at the square root – even more rapid transitions are possible given the dynamics.

The actual dynamics of scientific growth need further formalization. At the heart of the matter is the negative density dependence assumed by s-shaped curves. Restating this assumption, these models assume that publication growth will uniformly decline as more research is conducted. In ecology, where logistic growth modeling was first pioneered, there is an increasing recognition that, under some circumstances a population benefits from increased numbers. As a result, periods of very rapid or coalitional growth are possible. This phenomena is known as the Allee effect. The effect is most pronounced in social species such as birds, bees, fishes and herd animals.

Finding an equivalent phenomena in science and technology may require linking the numbers and careers of scientists to research output and specialization. Resource levels may play a role, as well as the minimum number of scientists required to maintain a productive research discipline. Maleszewska (2013) describes trend following behavior in the field of physics. Scientific research can be intensely social, and therefore it would not be surprising if the Allee effect were present in the sociology of science much as it is seen in ecology.

In the next section we report a simple empirical model which captures some of the salient features of the fold catastrophe.

## *An empirical model*

Data about word usage in general and keyword usage in particular is sparse and count-like in character. Normal approximations to the data, even using continuity corrections, fail when rates are less than five (Berenson et al., 1998). There are concerns with error-modeling, parsimony and predictive validity when using Gaussian models if these models are inappropriate. In particular, Gaussian models over-weight high count and high variance years. As a result critical information at the start of new publication trends is effectively discarded. Gaussian distributions, when used inappropriately, result in excess model parameters, and therefore an inability to generalize models when new data is present. Gaussian noise, when added to model structure, may result in negative predictions. Predictions of "negative" publication count are suspect both conceptually as well as validity-wise.

Unfortunately there are few existing Poisson models to be used as exemplars in the scientometrics, bibliometrics or informetrics literatures. Revolutionary work, dating back to 1976 and also due to de Solla Price, offers a probabilistic explanation of many cumulative advantage processes in science (de Solla Price, 1976). Much of the bibliometric work in probability borrows on an earlier tradition of biometrics (Bensman, 2005). Another exception is (Cunningham and Kwakkel, 2011), which presents a model for rapidly scanning or monitoring content areas of interest in order to trace dynamics and better predict future evolution.

Consider the Poisson distribution, a discrete distribution suitable for modeling count-like sources of publication data. The Poisson distribution is particularly suitable for modeling low frequency events as seen in emerging publication growth in new research fields. The probability density function of the Poisson distribution is as follows:

$$f(k, \lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{7}$$

In this function $\lambda$ is the Poisson rate, which may in turn be predicted by other independent variables. The variable $k$ takes on discrete quantities ($k = 0, 1, 2 \ldots n$) representing potential publication outputs. The resultant function gives the probability of $k$ publications occurring given a particular rate of publication output.

We now operationalize these dynamics in a simple time series model. The model is suitable for uncovering potential fold catastrophic dynamics in actual publication data. Our simple causal model of the Poisson rate $\lambda$ is solely a function of time ($t$), and three parameters.

$$\lambda = \exp(a + Hb(t - c)^{1/2}) \tag{8}$$

where $H$ is the Heaviside function and $H$ is 1 if $t > c$, or 0 otherwise.

Note that the rate parameter $\lambda$ is a function of time. Given an estimate of $\lambda$, shown as $\bar{\lambda}$, the likelihood of the model under this data and given this model parameter can be calculated. It is useful to report the logarithm of the likelihood, obtained by introducing the estimate to Eq. (7) above and taking the log-likelihood.

$$LL = k\log(\bar{\lambda}) - \bar{\lambda} - \log(\Gamma(k + 1)) \tag{9}$$

The presented model can be fitted to data using Poisson regression, analogous to the method used in Cunningham and Kwakkel (2011). The model predicts annual publication, not total or cumulative publication. The parameter $a$ is the baseline growth at the intercept year, parameter $b$ is associated with the equilibrium growth after the tipping point, and the parameter $c$ is associated with the bifurcation year. This generalizes the bifurcation phenomenon seen in Fig. 4.

We are aware of two other structural models, which is notable for widespread collection of data (Petersen et al., 2012). Petersen et al. also acknowledge periods of inflationary growth, but argue through statistical mechanics that these epochs are caused by periods high variation before standardization measures are taken. A further distinguishing feature is that Petersen et al. investigate general word usage in indexed books, while this paper investigates scientific articles. Cunningham and Kwakkel (2011) offer a forecasting model consistent with logistic growth in cumulative count of articles. This result is telling because the forecasted model consistently fails to reach saturation. Across all terms investigated the inflection point was near or at the end of the trend. In this context, although not a structural model, the work of Schmoch is anotable (Schmoch, 2007). Schmoch investigates the interwoven dynamics of scientific fields, scientific patenting, and technology sales. Particularly in patent counts Schmoch recognizes double-boom cycles caused by early research, then renewed interest in the underlying science during the commercialization phase. Schmoch's model is richly justified theoretically, but the paper does not offer an empirical model for trend forecasting.

## Application of the model

### *Case selection*

In order to assess the extent to which the postulated bifurcation dynamic plays a role across the sciences, we need a heterogeneous set of cases. That is, we need cases drawn from the full breadth and width of the sciences, instead of focusing on a single branch of science. Broadly selected cases eliminate the threat to validity that catastrophic change is limited in character, and that there is a case selection bias. That is, we believe the worst threat to validity for this type of research comes from the impression to "cherry pick" examples. Therefore, if in fact even a random sample demonstrates that rapid, inflationary growth predominates in scientific activity, we have strong evidence to generalize this claim. To this end, we analyzed a sample of 40,000 articles randomly sampled from the Web of Science for a particular year. We indexed titles, author provided keywords, and abstracts. We looked at phrases consisting of one word, two words, and three words. We excluded phrases containing stop words, based on a list of 333 stop words for the English language.

In order to find interesting phrases, we decided to not look for the most frequently occurring phrases. We identified the phrases that occurred between 130 and 150 times. This is a pragmatic choice intended to achieve a useful balance between frequently occurring words, which are inherently more representative of science, and narrowly specialized words. Narrowly specialized words use more technical language, adopting more precision in the meaning of the word, and therefore can be more readily mapped to activities at the research front in specific scientific subfields. This resulted in a list of 472 phrases. In order to identify a sample of cases that spans the sciences, we indexed the ISI subject categories and made a cross table of phrases by ISI subject categories. This cross table was analyzed using singular value decomposition and leading phrases by eigenvector are selected. This ensures a sampling of words across a multitude of disciplines. This results in 63 phrases, shown in Table 1. We choose 22 phrases from this list of 63, these are shown in gray in Table 1.

The search terms are exact and counts are reproducible by searches in the Web of Science database. Ideally there would be a functional equivalence between the search term, and a community of practicing researchers. A mismatch could occur because the term does not map to a functional concern in science, or because the term is polysemous, indicating different fields of study to different researchers. This could challenge a forecast by introducing additional conflicting signals in the underlying growth model. To further reduce the threat to validity posed by misspecification of our

**Table 1**
Identified phrases, the grayed phrases are selected for further analysis.

| Zro | Grafts | Integrin |
|---|---|---|
| Regimens | suppressor | Cyclin |
| Cisplatin | cterminal | Mcircle |
| Stents | Astrocytes | ism |
| Analog | Transistors | Allograft |
| Malignancies | nf | Cervical cancer |
| Self-assembly | Lasers | Transcription factors |
| Dwarf | Cytology | Receptor antagonist |
| rt | Diodes | Patients underwent |
| Malignancy | Novo | Cancer risk |
| Macrophage | Planetary | Mouse model |
| Confocal | Nanotube | Magnetic properties |
| Inducible | Epitaxial | Endothelial growth |
| Portal | cdc | Squamous cell |
| Pharmacokinetic | Oncology | Cell growth |
| Staging | xps | Published online |
| Transforming | Trafficking | Band gap |
| Adenosine | Disks | Solar cells |
| Mediates | gd | Photoelectron spectroscopy |
| Pathologic | es | Atomic force microscopy |
| Preclinical | Nanostructured | Vascular endothelial growth |

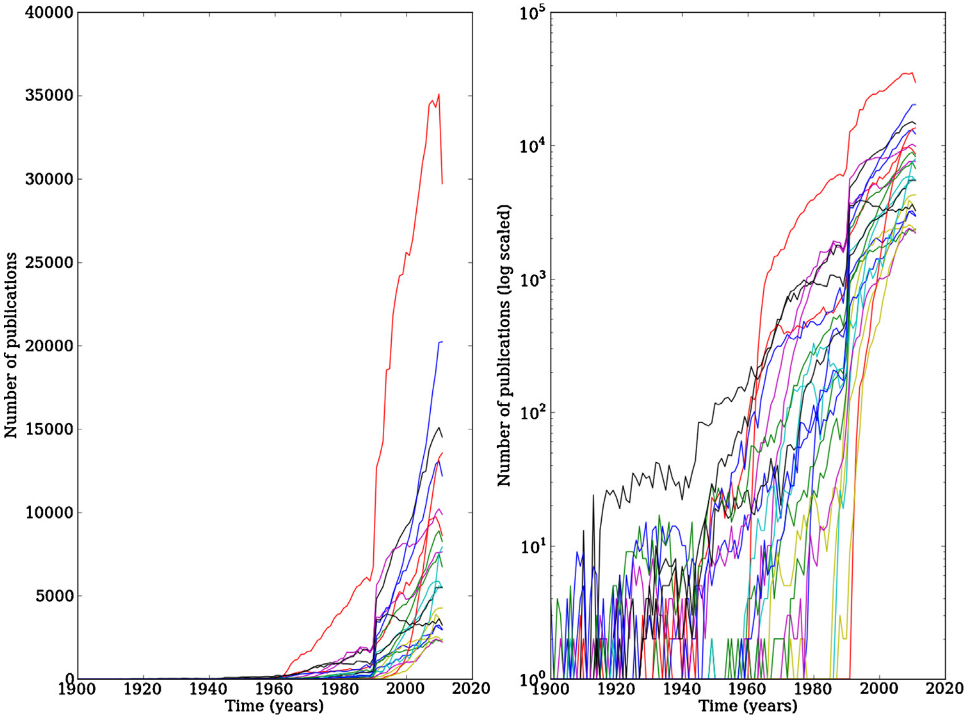


**Fig. 5.** Number of publications per year for each of the selected phrases.

terms, we have discussed them with various experts. The only term which raised concerns was "diodes." This term changed in meaning, reflecting radical changes in technology. We will shortly evaluate whether these trends are clear and unambiguous, and whether any model is sufficient for modeling the growth.A second concern would be raised if we were to use our forecasts to evaluate the health or promise the specific research communities involved. If care was not taken to comprehensively scope the research community through an appropriate selection of queries, incorrect inferences could be drawn concerning the community. We do not attempt such inferences in this paper, and are therefore correspondingly modest that we are measuring term usage and not the growth of scientific communities. Nonetheless we acknowledge results in sociology of science or scientometrics which are attempting such recommendations. Despite curtailing the goal of the modeling from scientific communities to scientific words, tracking word usage in science and the internet remains of intense interest (Petersen et al., 2012).Next, we queried the Web of Science for each of these 22 phrases under the topic field. We retrieved the number of publications per year for each of these 22 phrases (queried on January 10 2012). Resulting in time series data for each of these 22 phrases specifying the number of publications per year for each of these 22 phrases. Fig 5 shows the time series for each and a log scaled version of the same data. These results already suggest a rapid sudden growth for at least some of the keywords. It also shows that for the most recent years, the data is still incomplete. This is a known issue with ISI data. The data for 2011 is not complete yet, and the data even contains already some 2012 publications. For the remainder of the analysis, we restrict ourselves to data up to 2010.

*Fitting the model*

The model is implemented in the Python programming language (van Rossum, 1995). The optimization is performed using the PyEvolve genetic algorithm library (Perone, 2009). Matplotlib was used for the visualizations (Hunter, 2007).

**Table 2**
Results for fitting the bifurcation model to the 22 phrases.

| Keyword | Parameters of the model | | | | |
|---|---|---|---|---|---|
| | Intercept | Growth | Year of bifurcation | Growth of bifurcation | Log likelihood |
| Adenosine | 0.685053 | 0.003584 | 33 | 0.849211 | −6173.25 |
| Astrocytes | 0.465187 | 0.01227 | 70 | 0.993455 | −2599.59 |
| Atomic force microscopy | 0.208794 | 0.035534 | 86 | 0.989582 | −2786.12 |
| Cervical cancer | 0.017917 | 0.049523 | 84 | 0.509058 | −963.017 |
| Cisplatin | 0.880544 | 0.011849 | 70 | 0.961447 | −2390.65 |
| Diodes | 0.151526 | 0.075837 | 90 | 0.197996 | −3226.46 |
| Endothelial growth | 0.918966 | 0.028396 | 83 | 0.97932 | −2716.32 |
| Grafts | 0.861294 | 0.075368 | 83 | 0.122493 | −4403.63 |
| Integrin | 0.924806 | 0.019446 | 80 | 0.957764 | −4095.69 |
| Lasers | 0.750283 | 0.01806 | 41 | 0.961276 | −13,882.3 |
| Macrophage | 0.852808 | 0.008119 | 49 | 0.996081 | −9348.23 |
| Magnetic properties | 0.370629 | 0.073569 | 87 | 0.243715 | −3947.99 |
| Malignancy | 0.56042 | 0.052218 | 75 | 0.487983 | −3436.57 |
| Mouse model | 0.184304 | 0.044947 | 79 | 0.871613 | −5622.3 |
| Nanostructured | 0.133646 | 0.03584 | 94 | 0.99455 | −1021.06 |
| Nanotube | 0.073023 | 0.048622 | 94 | 0.993213 | −3426.08 |
| Pharmacokinetic | 0.723391 | 0.011218 | 54 | 0.9621 | −4893.23 |
| Photoelectron spectroscopy | 0.00633 | 0.027406 | 67 | 0.870907 | −3296.27 |
| Self-assembly | 0.032624 | 0.033073 | 88 | 0.998933 | −578.17 |
| Solar cells | 0.001602 | 0.027189 | 72 | 0.86059 | −2129.63 |
| Squamous cell | 0.91863 | 0.035228 | 75 | 0.664755 | −1839.59 |
| Stents | 0.189862 | 0.02437 | 83 | 0.961222 | −867.196 |

The model contains the following parameters

- A constant (constant)
- A growth rate (growth)
- The year at which the bifurcation takes place (bifurcate)
- The growth rate after the bifurcation (growth of bifurcate).

The constant, growth rate, and bifurcation growth rate are constrained between 0 and 1. The year of the bifurcation is unconstrained, meaning that the fitted model need not have a bifurcation over the period taken into consideration in the analysis. All the data runs from 1900 up to and including 2009. The maximum likelihood parameters for the model are calculated, using the parameters as listed above, and the log-likelihood equation given in Eq. (9), above.

The algorithm is ran for 250 generations, and each generation contains 2500 members. The crossover rate is set to 0.5 and mutation rate is set to 0.15. Table 2 shows the resulting parameterization and likelihood for each of the 22 phrases. Looking at the bifurcate column, we notice that the model always found a year of bifurcation in the data, this despite the fact that this parameter was unconstrained. Fig. 6 shows an illustration of the model. The original data is shown in dots, and the line is the model fit. The results suggest that the model fits the data well.

Fig. 5 shows one of the 22 models listed above, graphing the trend and the fit in the word "self-assembly."

It is useful to further examine the parameters in the 24 models given above. This is displayed in a matrix scatter plot, shown in Fig. 7, below. The model suggests our selected terms, although selected at random, were derived from two populations – one which was still very new in the year 1900, and another which was already quite mature in the year 1900. Growth is quite tightly constrained across a limited growth range. Bifurcations have apparently grown more frequent over time. When bifurcations do occur, very dramatic growth is the most frequent outcome. The scatter plots describe other potential cross-correlations between parameters. Terms which are introduced later into the scientific vocabulary may show more rapid growth and diffusion. The more the initial publication output in 1900, the less time before bifurcation. Higher baseline growth is generally associated with a

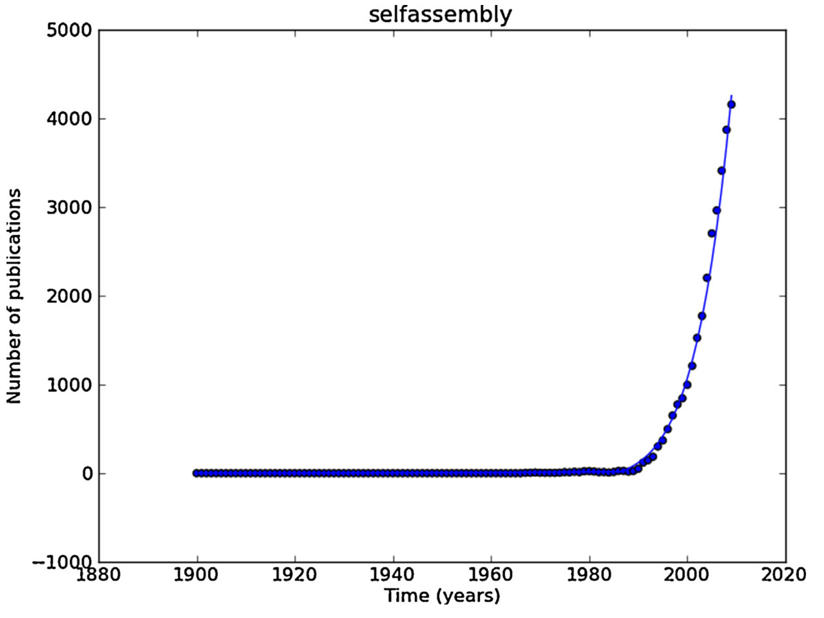


**Fig. 6.** An illustration of the fit of the model for the phrase 'self-assembly'The dots are the original data, the line is the fitted model.

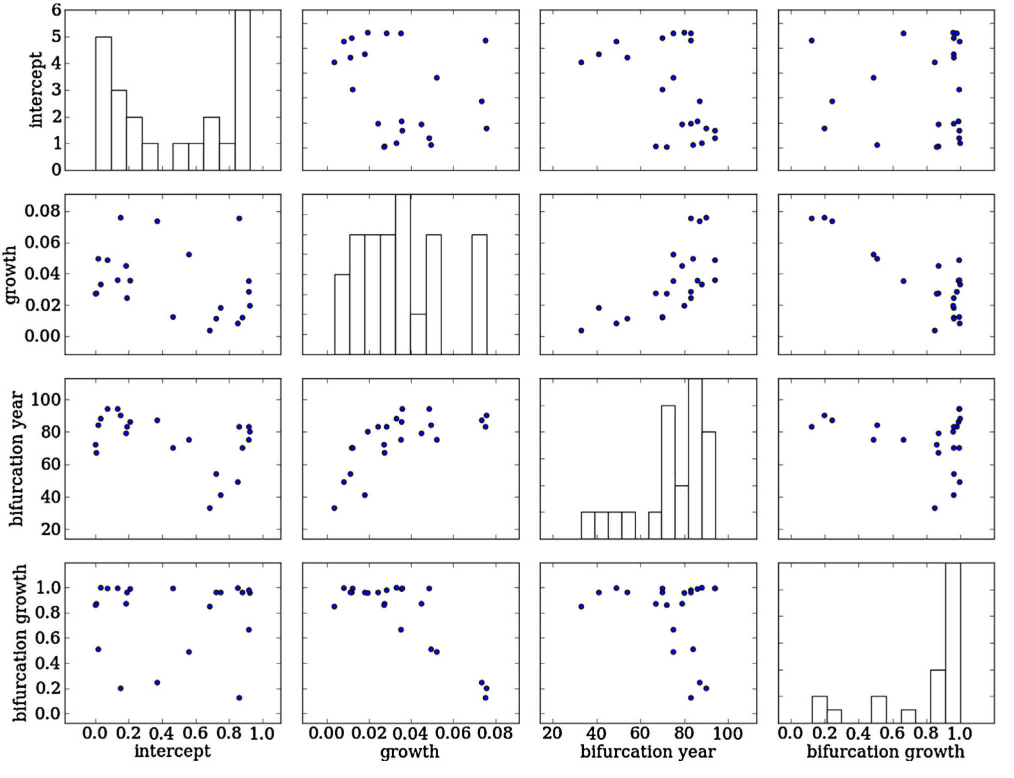


**Fig. 7.** Matrix scatter plot of model parameters.

later year for bifurcation, and slower growth after the bifurcation. The later the bifurcation year, generally the slower the resultant growth after bifurcation. These results are only descriptive, and drawn from a limited sample of models, but they are nonetheless suggestive.

*Comparison with the Bass model*

Having shown that the bifurcation model can be fitted reasonably well to data pertaining to word usage, the next question is whether the model offers a better explanation then some of the other possible models. Many models can in principle be used to describe the diffusion of word usage through a community of scientists, including the Fisher-Pry, Pearl, Gompertz and Bass models (Bass, 1969; Porter et al., 1991; Roper et al., 2011). These models differ in their postulated underlying non-linear processes of diffusion and saturation. The underlying process is often justified using dynamic models of population growth.

For instance, the Bass model was originally intended to describe the first adoption of a new technology. The model assumes that there are two key adoption processes. Some fraction of consumers will unconditionally adopt a new technology at a given period of time. This fraction is known as the "coefficient of internal innovation". Another fraction of consumers will adopt a new technology only if their peers have adopted the technology. This fraction is known as the "coefficient of external innovation" Together the internal and external coefficient determine the ultimate speed and extent of technology adoption. Since the model describes the first adoption only, once all prospective customers have adopted the technology predicted new adoptions then cease. The Bass model is closely related to the Pearl and Fisher-Pry models of technology adoption. The difference is that the Bass model predicts the rate of new adoption, while the Pearl and Fisher-Pry models forecast the cumulative new adoptions. The Bass model is given below.

$$A(t) = \frac{M - Me^{-(p+q)(t+t_0)}}{1 + \frac{q}{p}e^{-(p+q)(t+t_0)}}$$

Here $M$ is the size of the potential users, $p$ is the coefficient of innovation, $q$ is the coefficient of imitation, and $t_0$ is the year in which the product appeared on the market. We fitted the Bass model to

**Table 3**
Results for fitting the Bass model to the 22 phrases.

| Keyword | Parameters of the bass model | | | | |
|---|---|---|---|---|---|
| | $p$ | $q$ | $m$ | $t_0$ | Log likelihood |
| Adenosine | 0.000118 | 0.051935 | 4373 | 19 | −17,196.2 |
| Astrocytes | 0.000484 | 0.040808 | 1339 | 12 | −22,654.3 |
| Atomic force microscopy | 0.000875 | 0.034507 | 2221 | 8 | −58,905.2 |
| Cervical cancer | 0.00041 | 0.043822 | 1055 | 18 | −21,784.1 |
| Cisplatin | 0.0009 | 0.039014 | 1632 | 1 | −27,101.4 |
| Diodes | 0.000557 | 0.002909 | 2184 | 527 | −165,128 |
| Endothelial growth | 0.000177 | 0.044728 | 5791 | 7 | −49,583.3 |
| Grafts | 0.00107 | 0.025936 | 7134 | 24 | −134,336 |
| Integrin | 0.000687 | 0.032938 | 1410 | 18 | −33,556.6 |
| Lasers | 0.001141 | 0.010884 | 17,452 | 90 | −537,632 |
| macrophage | 0.001125 | 0.017741 | 3356 | 88 | −170,938 |
| Magnetic properties | 0.001035 | 0.035457 | 6023 | 3 | −94,582.6 |
| Malignancy | 0.000496 | 0.039478 | 4957 | 8 | −58,251.7 |
| Mouse model | 0.000421 | 0.044673 | 9075 | 1 | −126,193 |
| Nanostructured | 0.000398 | 0.044137 | 1043 | 1 | −21,636.1 |
| Nanotube | 0.000639 | 0.03134 | 3286 | 13 | −95,054.5 |
| Pharmacokinetic | 0.000253 | 0.01797 | 3536 | 132 | −105,080 |
| Photoelectron spectroscopy | 0.000802 | 0.037507 | 2625 | 4 | −43,602.3 |
| Self-assembly | 0.000275 | 0.039636 | 1729 | 15 | −28,446.8 |
| Solar cells | 0.000893 | 0.037394 | 1617 | 2 | −29,322 |
| Squamous cell | 0.000473 | 0.041792 | 3334 | 3 | −33,507.7 |
| Stents | 0.000775 | 0.030308 | 778 | 24 | −22,366.3 |

**Table 4**
Comparison of quality of fit of the bifurcation model and the Bass model.

| Keyword | Fold model | Bass model | Log-odds |
|---|---|---|---|
| Cisplatin | −2390.65 | −27,101.4 | 24,710.71959 |
| Astrocytes | −2599.59 | −22,654.3 | 20,054.66584 |
| Lasers | −13,882.3 | −537,632 | 523,749.5359 |
| Photoelectron spectroscopy | −3296.27 | −43,602.3 | 40,306.06975 |
| Pharmacokinetic | −4893.23 | −105,080 | 100,186.722 |
| Integrin | −4095.69 | −33,556.6 | 29,460.94394 |
| Grafts | −4403.63 | −134,336 | 129,931.9136 |
| Cervical cancer | −963.017 | −21,784.1 | 20,821.05962 |
| Endothelial growth | −2716.32 | −49,583.3 | 46,867.02369 |
| Diodes | −3226.46 | −165,128 | 161,901.5086 |
| Atomic force microscopy | −2786.12 | −58,905.2 | 56,119.04462 |
| Macrophage | −9348.23 | −170,938 | 161,590.0407 |
| Selfassembly | −578.17 | −28,446.8 | 27,868.59109 |
| Squamous cell | −1839.59 | −33,507.7 | 31,668.12734 |
| Magnetic properties | −3947.99 | −94,582.6 | 90,634.6568 |
| Malignancy | −3436.57 | −58,251.7 | 54,815.16083 |
| Nanotube | −3426.08 | −95,054.5 | 91,628.40046 |
| Solar cells | −2129.63 | −29,322 | 27,192.34149 |
| Stents | −867.196 | −22,366.3 | 21,499.07411 |
| Nanostructured | −1021.06 | −21,636.1 | 20,615.0359 |
| Adenosine | −6173.25 | −17,196.2 | 11,022.94444 |
| Mouse model | −5622.3 | −126,193 | 120,570.4395 |

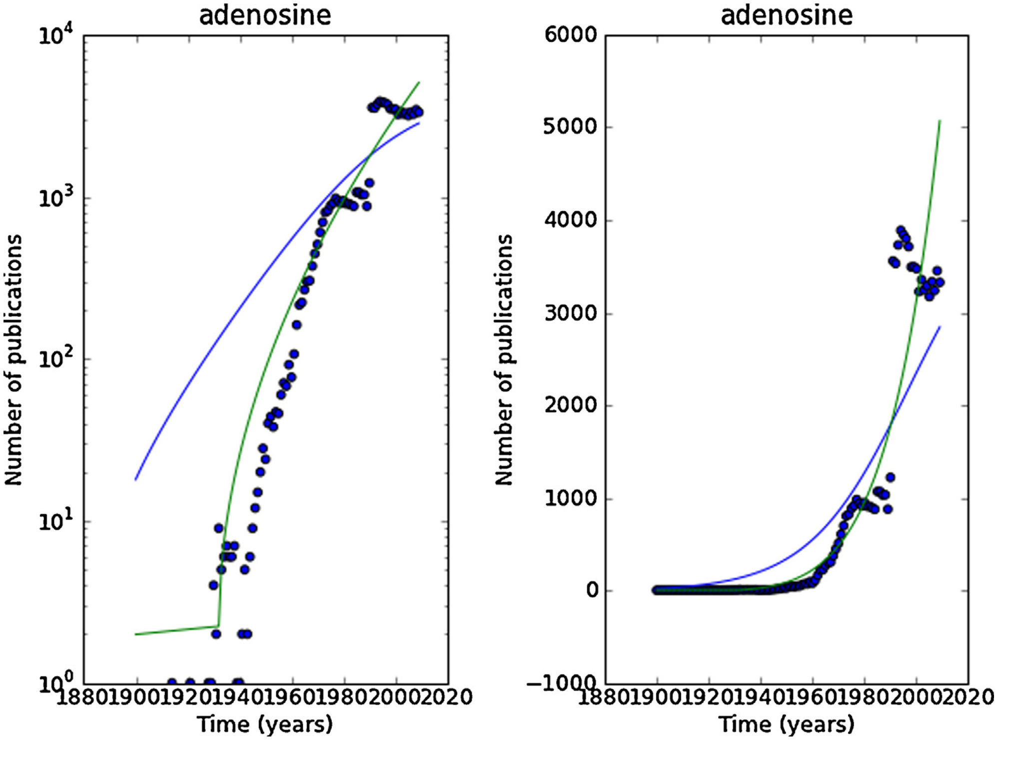


**Fig. 8.** Comparison of the Bass model and the Bifurcation model. Green is the bifurcation model and blue is the Bass model. Again, the dots are the original data.

word use data using poison regression. We choose to compare the bifurcation model to the Bass model for two reasons. First, the Bass model represents the state of the art in modeling diffusion processes in science.

Second, the Bass model has the same number of parameters, thus allowing for a fair comparison. A fundamental problem in model selection is how to compare two models with a different number of free parameters. In this case, merely comparing on log likelihood skews the comparison in favor of the model with more free parameters. There exists a range of metrics and associated motivating ides to support in this process.

One such metric of quality is Akaike's Information Criteria (AIC) (Akaike, 1974). This metric rewards high likelihood models, while penalizing models according to the number of free parameters assumed by the model structure. The AIC and related metrics bear more than a passing resemblance to Occam's razor – a heuristic which has served science well for many centuries. Our confidence in model results is conditioned on the correctness of the structural explanation of the data as offered by the model. Metrics like AIC supports the evaluation of competing explanations of the data, allowing us to simultaneously evaluate both the structural and uncertainty issues with the model. However, since in our case, both the Bass model and our model have the same number of free parameters, a comparison of the log likelihood directly can be used. It will reveal which of the two models offers a better structural explanation of the data.

For fitting the model, we again used a genetic algorithm. The algorithm is ran for 250 generations, and each generation contains 2500 members. The crossover rate is set to 0.5 and mutation rate is set to 0.15. Table 3 shows the resulting parameterization and likelihood for each of the 22 phrases.

In order to compare the quality of fit of both models we can use the log likelihood directly since both models have the same number of parameters. This results in Table 4, which shows univocally that the bifurcation model offers a significantly better structural explanation of the data than the Bass model. Fig. 8 depicts this visually for the phrase where the difference in quality of fit is the smallest. As can be seen for this randomly selected case, the very rapid early growth is much faster than can be explained by the Bass model, but fits very well with our bifurcation model. In general, the lower

**Table 5**
Results for fitting the bifurcation model with two additional dummy variables to the 22 phrases.

| Keyword | Parameters of the bifurcation model with dummies | | | | | | |
|---|---|---|---|---|---|---|---|
| | Constant | Growth | Bifurcate | Growth of bifurcate | Dummy constant | Dummy growth | Log likelihood |
| Cisplatin | 0.669689 | 0.023483 | 67 | 0.650318 | 0.139093 | 0.005006 | −2095.61 |
| Astrocytes | 0.46 | 0.001778 | 44 | 0.555268 | 0.47957 | 0.020433 | −1172.11 |
| Lasers | 0.232106 | 0.010484 | 26 | 0.927189 | 0.444263 | 0.002604 | −13,647.3 |
| Photoelectron spectroscopy | 0.352981 | 0.029046 | 60 | 0.442384 | 0.459763 | 0.01368 | −956.291 |
| Pharmacokinetic | 0.724355 | 0.053301 | 55 | 0.289834 | 0.164825 | 0.002581 | −6579.39 |
| Integrin | 0.062998 | 0.024247 | 161 | 0.049241 | 0.931699 | 0.039487 | −1409.6 |
| Grafts | 0.928309 | 0.030451 | 26 | 0.479524 | 0.422747 | 0.005058 | −1398.49 |
| Cervical cancer | 0.143111 | 0.040673 | 75 | 0.405714 | 6.08E−01 | 0.003052 | −503.454 |
| Endothelial growth | 0.25305 | 0.015876 | 75 | 0.720497 | 9.59E−01 | 0.015905 | −709.278 |
| Diodes | 0.356129 | 0.022232 | 43 | 0.664294 | 0.286316 | 0.005951 | −1753.73 |
| Atomic force microscopy | 0.058199 | 0.011879 | 89 | 0.605017 | 0.386769 | 0.039718 | −1155.76 |
| Macrophage | 0.597397 | 0.03529 | 28 | 0.433946 | 0.323006 | 0.00724 | −9688.33 |
| Selfassembly | 0.994223 | 0.007885 | 89 | 0.995321 | 0.480642 | 0.012952 | −632.723 |
| Squamous cell | 0.543567 | 0.006755 | 54 | 0.783854 | 0.563583 | 0.007954 | −722.89 |
| Magnetic properties | 1.74E−01 | 0.073564 | 153 | 0.258647 | 0.636708 | 0.005565 | −1115.13 |
| Malignancy | 0.225032 | 0.041286 | 48 | 0.364015 | 0.558141 | 0.007907 | −736.465 |
| Nanotube | 0.4825 | 0.01196 | 99 | 0.982393 | 7.13E−01 | 0.034207 | −1237.57 |
| Solar cells | 0.655114 | 0.04973 | 104 | 0.413614 | 2.64E−01 | 0.011174 | −2490.93 |
| Stents | 0.304107 | 0.016622 | 79 | 0.649476 | 0.058132 | 0.017599 | −510.901 |
| Nanostructured | 0.022005 | 0.002019 | 98 | 0.800522 | 0.354788 | 0.0437 | −538.148 |
| Adenosine | 6.98E−01 | 0.002418 | 27 | 0.761 | 0.211643 | 0.004718 | −6004.11 |
| Mouse model | 6.98E−01 | 0.003683 | 70 | 0.97089 | 5.92E−01 | 0.018443 | −532.479 |

likelihood and poorer fit for the Bass model derives from two sources. The Bass model consistently over-estimates the rate of early growth of term usage, and consistently under-estimates the potential for long-term growth in term usage. Further, the Bass model consistently estimates that the inflection of the curve is at the end of the full and observed trend of data. All three observations are consistent with modeling artifacts created as the Bass curve attempts to deal with short periods of inflationary growth.

*Indexing artifact*

Next to comparing the bifurcation model to other models for diffusion processes, another threat to validity of the model is that the results we see are entirely due to indexing artifacts. As is well known, ISI data is not stable and the indexing has changed over time. One very important change is the year ISI started to index abstracts. Fig. 8 already highlights this, for there is a strange jump around 1990. ISI started to index abstracts in 1991. The indexing of abstracts affects the results of our topic query, for from 1991 onward, the query is not only based on titles and author provided keywords, but also on abstracts, thus increasing the potential number of papers that are returned for a particular query topic. This indexing artifact can affect both the constant as well as the rate of growth. Thus, we extend to be model with two dummy variables, one for the constant and one for the growth. The dummies are set to zero prior to 1991. Model fitting is done identical to the foregoing. Thus, we use Poisson regression, and use a genetic algorithm. The algorithm is ran for 250 generations, and each generation contains 2500 members. The crossover rate is set to 0.5 and mutation rate is set to 0.15. Table 5 shows the results.

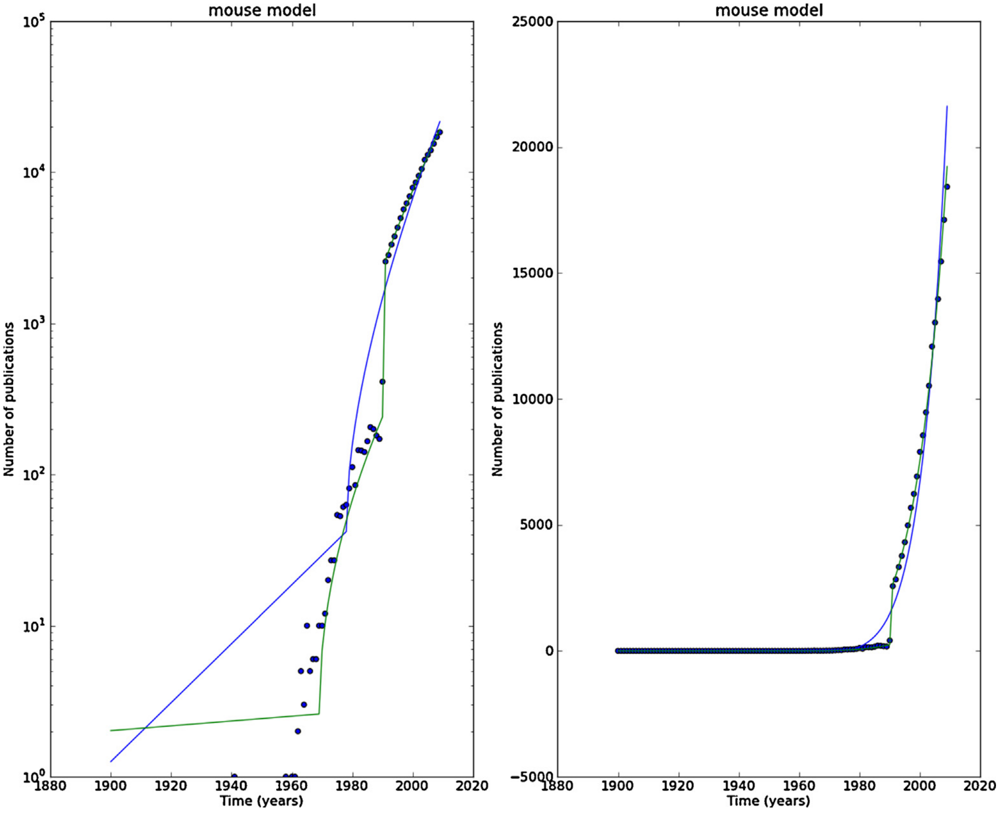


**Fig. 9.** Comparison of the bifurcation model and the bifurcation model with dummy. Green is the bifurcation model with dummy and blue is the normal bifurcation model. Again, the dots are the original data.

Fig. 9 shows visually the impact of adding the dummy to the data. From the log plot on the left, we deduce that the year of bifurcation in the model with dummy variables is a few years earlier, then in the original model. However, these results in Fig. 9 also clear show that the 1991 indexing artifact does not invalidate the fact that there has been a tipping point after which the rate of growth was substantially faster. Twenty out of twenty-two cases show a bifurcation even in the presence of the 1991 dummy.

Looking at Table 5, we observe that only for 'integrin' and 'magnetic properties', the model gives a bifurcation year outside of the range of the model. Fig. 10 shows the 'integrin' case. It appears that the number of articles using this keyword started to grow rapidly just prior to 1991. In fact, the first article on integrin appeared in 1973, followed by another one in 1976 and one in 1977. In 1986, four articles appear. In 1990, so prior to the indexing of abstracts, this had already risen to 147. So in 5 years, the number of articles had increased more than 35 times. This suggests that there very well could have been a tipping point for integrin, but that it is partly obscured by the fact that this point almost coincided with the change in indexing. Moreover, looking at Fig. 10, it appears that the growth just prior to 1991 is in fact faster than postulated by our bifurcation model. We return to this point in the discussion below.

The starting point for this paper was that the logistic growth that is often postulated to apply to science ever since the work of de Sola Price is problematic for both theoretical and empirical reasons. As we have shown, growth substantially faster than exponential is in fact to be found across the sciences. The randomly selected keywords from across the breadth of the sciences show for all but a few cases fold like dynamics, which is at odds with the postulated exponential or logistic growth model. That is, we have provided systematic empirical evidence, on top of the more anecdotal

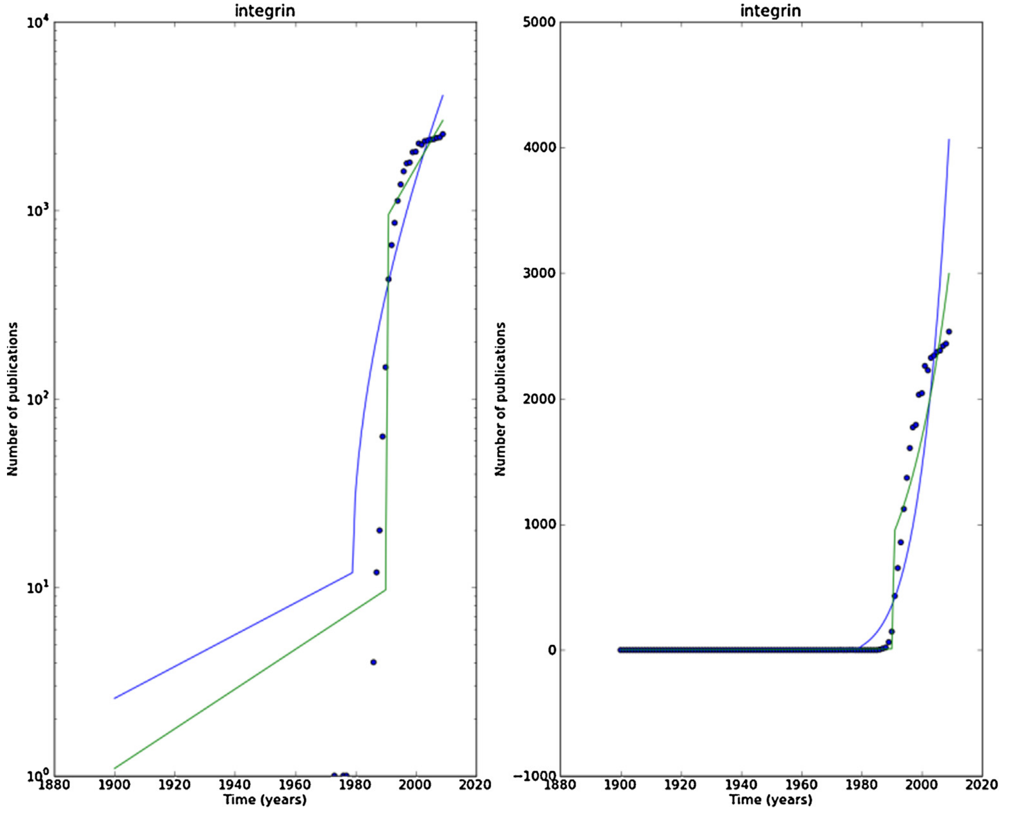


**Fig. 10.** Comparison of the bifurcation model and the bifurcation model with dummy for 'integrin'. Green is the bifurcation model with dummy and blue is the normal bifurcation model. Again, the dots are the original data.

evidence offered in (Cunningham and Kwakkel, 2011), that the dynamics of scientific publication are in fact rive with sudden burst of activity.

## Discussion

The idea of a tipping point in the sciences implies that the rate of publication on a particular topic can either rapidly increase, or even collapse. All the cases that have been shown in the foregoing show a rapid increase only. Collapse dynamics are not present. This arguably is a consequence of the sampling strategy that we utilized. We looked at phrases that have a relatively high usage in a random sample of ISI data. We identified a set of phrases that are somewhere in the middle in terms of their frequency, and then selected from this set a subset that spans the sciences. Phrases denoting topics that have already collapsed would not show up in this middle tier and thus have not been selected. A random sample, taken at some earlier year, say in the early nineties, might help in identifying phrases that show a collapse.

A second issue is that the model we used postulates a single bifurcation. It is conceivable, however, that a particular topic bifurcates more than once. That is, it bifurcates at some point in the past, for example after its initial discovery. This bifurcation is followed by a rapid growth that stabilized at some point. In Kuhnian terms, this process of stabilization can be perceived as normal science, or the mopping up after the profound discovery or development has done its work. However, it is conceivable that during this stabilization phase new discoveries or breakthroughs are made related to the topic, thus sparking another round of frantic research on the topic. Fig. 11 shows a possible case of this double bifurcation dynamic for diodes. To show that the second growth cannot be explained

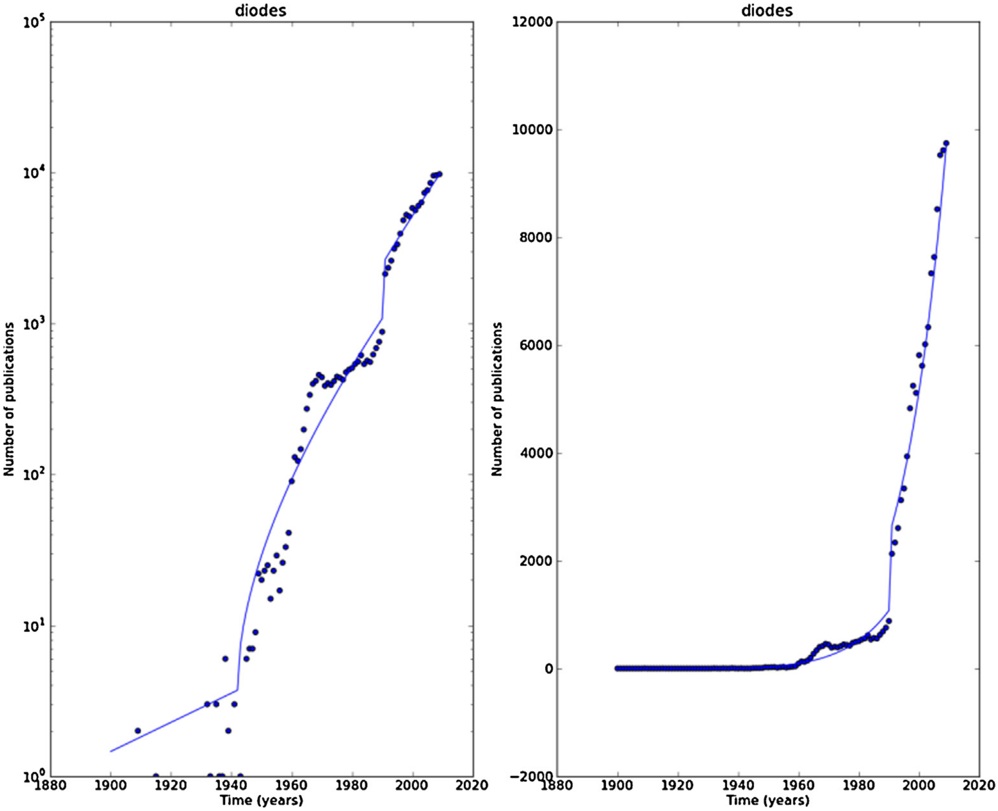


**Fig. 11.** Potential double bifurcation for 'diodes'. Blue is the bifurcation model with dummy. The dots are the original data.

solely by the 1991 indexing artifact, we fitted the dummy model to the data. As can be seen, the rapid second growth of the topic started a few years prior to 1991.

A third issue is that the current model postulates that the growth after the bifurcation is squared. If one analyzes the results in detail, with a particular focus to the visualization of the fitted model, it appears that the growth after the bifurcation in some cases might be faster than squared. A possible simple extension to the model would be to turn the root into a free parameter in the optimization. This would allow for the investigation of the growth after the bifurcation across the sciences.

## Conclusion

The starting point of this paper was the claim that scientific knowledge can demonstrate explosive growth in short periods of time. This explosive growth is significantly faster than the kind of growth one would expect based on various existing diffusion modes, such as the Bass or Gompertz model. These models cannot account for the discontinuous and sudden growth that is observed in the real world. In one notable example, the field of engineering and technology management grew more rapidly in the 4 years after 1980 than it was expected to grow for the next 40 years (Cunningham and Kwakkel, 2011). This paper adds another set of cases to this that all show explosive growth well beyond what would be expected according to the Bass model.

In light of the empirical observation of rapid growth, we formulated a model based on the idea of folds from mathematical catastrophe theory. This model postulates a slow growth process leading up to a point at which the growth curve suddenly changes dramatically. This point at which the growth shows a sudden discontinuous jump is more popularly known as a tipping points. The presented model was fitted to data using Poisson regression. We compared the model with two rival explanations of the data. The first alternative explanation came in the form of the Bass model. Here, the difference in the quality of fit of the model clearly demonstrated that the bifurcation model can explain the data much better than the Bass model. The second alternative explanation was indexing artifacts. To assess the impact of the single biggest indexing artifact in ISI data, the indexing of abstracts in 1991, we extended the model and included two dummy variables. When we fitted this modified model to the data, we were still able to find a bifurcation in all but two cases. Thus suggesting that the rapid growth cannot be explained by indexing artifacts alone.

The extent to which the model can be used to predict future tipping points or bifurcations is limited, for the model only postulates the presence of a tipping point. It does not open up how or why a tipping point occurs. However, despite not offering an explanation for the bifurcation, the model still can be used in the context of monitoring scientific developments, and can provide early warning that a particular word or phrase has passed a tipping point. This insight can be used for various purposes including individual scientists updating their vocabulary, funding agencies tailoring research calls, and technology management. This research also contributes to the emerging subfield of innovation forecasting, where dynamical trends in science and patenting are examined for leading indicators of emerging fields of science and technology. Further research, however, is necessary to identify early warnings, or signal, of a potential bifurcation. Possible directions for such research include the assessment of the dispersion and variance of the data as compared to the model. Another direction of research is to tie this scientific bifurcation dynamic to the wider avenue of research in the complexity sciences. Here, there is active research into early signals of tipping points in all sorts of systems. There is some evidence that systems close to a bifurcation point are slower to return to equilibrium if perturbed.

Dynamic theories of knowledge generation are of intrinsic interest to a number of different fields of research. The theory of dynamic capabilities is used by practitioners in strategic and technology management; these researchers are interested in how firms acquire and broker knowledge for economic advantage (Malerba, 2002; Teece et al., 1997). Direct knowledge acquisition from universities to industry occurs rarely, yet it may constitute the primary source of fundamental sources of knowledge in firms (Zucker et al., 2002). This theory within the strategic management literature draws upon evolutionary economic approaches (Zollo and Winter, 2002). The community of practice literature examines the technological and social strategies adopted by networks of experts attempting to remain productive while dealing with rapid changes in knowledge (Fuhr and Fuchs-Kittowski, 2004). The presented findings on the omnipresence of fold dynamics across the sciences have ramifications for each of these fields. Topics in science can grow much faster than the exponential and logistic model postulate, severely reducing the reaction time when monitoring these developments. Moreover, these discontinuous jumps means that there will be surprises or black swans in future scientific growth.

## References


Akaike, H., 1974. A new look at the statistical model identification. IEEE Transaction on Automatic Control 19 (6) 716–723.

Arnold, V.I., 1984. Catastrophe Theory. Springer-Verlag, Berlin, Germany.

Arnold, V.I., Afrajmovich, V.S., Il'yashenko, Y.S., Shil'nikov, L.P., 1994. Dynamical Systems V, Encyclopaedia of Mathematical Sciences. Springer-Verlag, Berlin, Germany.

Bass, F., 1969. A new product growth model for consumer durables. Management Science 15 (5) 215–227.

Bengisu, M., Nekhili, R., 2006. Forecasting emerging technologies with the aid of science and technology databases. Technological Forecasting and Social Change 73 (7) 835–844.

Bensman, S.J., 2005. Urquhart and probability: the transition from librarianship to library and information science. Journal of the American Society for Information Science 56 (2) 189–214.

Berenson, M.L., Levine, D.M., Tucker, T., Anderson, N., 1998. The Normal Approximation to the Binomial and Poisson Distributions, Business Statistics. Prentice-Hall Inc., Upper Saddle River.

Bettencourt, L.M.A., Lobo, J., Helbing, D., Kuhnert, C., West, G.B., 2007. Growth, innovation, scaling, and the pace of life in cities. Proceedings of the National Academy of Science of the United States of America 104 (17) 7301–7306.

Boyack, K.W., Borner, K., Klavans, R., 2007. Mapping the evolution and structure of chemistry research Proceedings of ISSI 2007: 11th International Conference of the International Society for Scientometrics and Informetrics, vol. 1. , pp. 112–212.

Cunningham, S., Kwakkel, J.H., 2011. Innovation forecasting: a case study of the management of engineering and technology literature. Technological Forecasting and Social Change 78 (2) 346–357.

Daim, T.U., Rueda, G., Martin, H., Gerdsri, P., 2006. Forecasting emerging technologies: use of bibliometrics and patent analysis. Technological Forecasting and Social Change 73 (8) 981–1012.

de Solla Price, D., 1961. Science Since Babylon. Yale University Press, New Haven, CT.

de Solla Price, D., 1963. Little Science, Big Science. Columbia University Press, New York.

de Solla Price, D., 1976. A general theory of bibliometric and other cumulative advantage processes. Journal of the American Society for Information Science 27 (5-6) 292–306.

de Solla Price, D., 1983. This week's citation classic: little science, big science. Current Contents 29, 18.

Fisher, J.C., Pry, R.H., 1971. A simple substitution model of technological change. Technological Forecasting and Social Change 3, 75–88.

Fuhr, D., Fuchs-Kittowski, F., 2004. Against hierarchy and chaos: knowledge coproduction in nets of experts. Journal of Universal Computer Science 10 (3) 176–185.

Garfield, E., 2004. Historiographic mapping of knowledge domains literature. Journal of Information Science 30 (2) 119–145.

Gladwell, M., 2002. The Tipping Point: How Little Things Can Make a Big Difference. First Back Bay, Boston.

Guckenheimer, J., Haiduc, R., 2003. Canards at folded nodes. Moscow Mathematical Journal 5 (1) 91–103.

Han, C.-S., Lee, S.K., England, M., 2010. Transition to postmodern science: related scientometric data. Scientometrics 84, 391–401.

Hicks, D., Katz, J.S., 1996. Where is science going? Science Technology Human Values 21, 379–406.

Hubbell, S.P., 2001. The Unified Neutral Theory of Biodiversity and Biogeography. Princeton University Press, Princeton.

Hunter, J.D., 2007. Matplotlib: a 2D graphics environment. Computing in Science and Engineering 9 (3) 90–95.

Lariviere, V., Archambault, E., Gingras, Y., 2008. Long-term variations in the aging of scientific literature. From exponential growth to steady-state science (1900–2004). Journal of the American Society for Information Science and Technology 59, 288–296.

Malerba, F., 2002. Sectoral systems of innovation and production. Research Policy 31, 247–264.

Maleszewska, M., 2013. Fashion rules in physics. Nature 500, 113.

Mendez, V., Sans, C., Llopis, I., Campos, D., 2011. Extinction conditions for isolated populations with Allee effect. Mathematical Biosciences 232 (1) 78–86.

Merton, R.K., 1960. The Matthew effect in science. Science 159 (3810) 56–63.

Miranda, L.C.M., Lima, C.A.S., 2010. On the logistic modeling and forecasting of evolutionary processes: application to human population dynamics. Technological Forecasting and Social Change 77 (5) 699–711.

Mitchell, M., 2009. Complexity, A Guided Tour. Oxford University Press, Oxford.

Noyons, E.C.M., van Raan, A.F.J., 1988. Monitoring scientific developments from a dynamic perspective: self-organized structure to map neural network research. Journal of the American Society for Information Science 49 (1) 61–68.

Pearl, R., Reed, L.J., 1920. On the rate of growth of the population of the United States since 1790 and its mathematical representation. Proceedings of the National Academy of Sciences 6, 275–288.

Perone, C.S., 2009. Pyevolve: a python open-source framework for genetic algorithms. ACM SIGEVOlution 4 (1) 12–20.

Petersen, A.M., Tenenbaum, J., Havlin, S., Stanley, H.E., 2012. Statistical laws governing fluctuations in word use from birth to death. Scientific Reports 2, 313.

Porter, A.L., Roper, A.T., Mason, T.W., Rossini, F.A., Banks, J., 1991. Forecasting and Management of Technology. John Wiley & Sons, New York.

Reid, E.O., 1997. Evolution of a body of knowledge: an analysis of terrorism research. Information Processing & Management 33 (1) 91–106.

Rip, A., Courtial, J.P., 1984. Co-word maps of biotechnology: an example of cognitive scientometrics. Scientometrics 6, 265–269.

Roper, A.T., Cunningham, S.W., Porter, A.L., Mason, T.W., Rossini, F.A., Banks, J., 2011. Forecasting and Management of Technology. John Wiley & Sons, Hoboken, New Jersey.
Schmoch, U., 2007. Double-boom cycles and the comeback of science-push and market-pull. Research Policy 36 (7) 1000–1015.
Teece, D.J., Pisano, G.P., Shuen, A., 1997. Dynamic capabilities and strategic management. Strategic Management Journal 18 (7) 509–533.
van den Besselaar, P., Heimeriks, G., 2006. Mapping research topics using word-reference co-occurrences: a method and an exploratory case study. Scientometrics 68 (3) 377–393.
van Rossum, G., 1995. Python Reference Manual. CWI.
Verhulst, P.-F., 1838. Notice sur la loi que la population poursuit dans son accroissement. Correspondance Mathématique et Physique 10, 113–121.
von Foerster, H., Mora, P.L.A., 1960. Doomsday: Friday, 13 November, A.D. 2026. Science 132, 1291–1295.
Zitt, M., 1991. A simple method for dynamic scientometrics using lexical analysis. Scientometrics 22 (1) 229–252.
Zollo, M., Winter, S.G., 2002. Deliberate learning and the evolution of dynamic capabilities. Organization Science 13 (3) 339–351.
Zucker, L.G., Darby, M.R., Amstrong, J.S., 2002. Commercializing knowledge: university science, knowledge capture, and firm performance in biotechnology. Management Science 48 (1) 138–153.