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

Scientometrics, Vol. 51, No. 1 (2001) 9–36

Application of statistical physics methods and concepts to the study of science & technology systems

L. A. N. AMARAL,[1] P. GOPIKRISHNAN,[1] K. MATIA,[1] V. PLEROU,[1,2] H. E. STANLEY[1]

[1] *Center for Polymer Studies and Department of Physics, Boston University, Boston, MA (USA)*
[2] *Department of Physics, Boston College, Boston, MA (USA)*

We apply methods and concepts of statistical physics to the study of science & technology (S&T) systems. Specifically, our research is motivated by two concepts of fundamental importance in modern statistical physics: scaling and universality. We try to identify robust, universal, characteristics of the evolution of S&T systems that can provide guidance to forecasting the impact of changes in funding. We quantify the production of research in a novel fashion inspired by our previous study of the growth dynamics of business firms. We study the production of research from the point of view both of inputs (R&D funding) and of outputs (publications and patents) and find the existence of scaling laws describing the growth of these quantities. We also analyze R&D systems of different countries to test the "universality" of our results. We hypothesize that the proposed methods may be particularly useful for fields of S&T (or for levels of aggregation) for which either not enough information is available, or for which evolution is so fast that there is not enough time to collect enough data to make an informed decision.

Introduction

The choices of decision makers have a fundamental role in the development and evolution of a national science and technology (S&T) system. This fact is clearly highlighted by the impact, nearly half a century ago, of Vannebar Bush's "Report to the President on a Program for Postwar Scientific Research" on the establishment of the bipartisan approach of the USA to research and development (R&D).[1,2]

Decisions on how much to spend on R&D, how to partition funds among disciplines (e.g., life sciences or natural sciences?), or how much to spend on focussed problems (e.g., Human Genome Project, global warming, renewable sources of energy) can have a dramatic impact of what advances might occur first, and may even seriously jeopardize the competitiveness of a S&T system if a wrong strategic decision is made.[3,4] Such concerns are particularly pressing nowadays due to (i) the increased scale and resources of the S&T systems as compared to 50 years ago, (ii) the increased rate of change of

0138–9130/2001/US $ 15.00

scientific advances, and (iii) the multidisciplinary character of cutting-edge research (consider, e.g., the new field of bioinformatics, where biologists, mathematicians and physicists are sometimes cooperating and sometimes competing).

To make informed choices, the decision maker needs information that is timely, reliable, and clear. In an answer to these needs, the field of quantitative S&T studies has gone through a revolutionary period,[5] with many developments occurring in the identification of new indicators.[6] In spite of these new advances this is still an extremely complex problem, for instance, indicators are by definition retrospective and heuristic.[1] Moreover, there are many difficulties in developing indicators[7] that are general and robust and can be applied across (i) the different S&T fields, (ii) for different aggregation[8–12] levels (from research groups to entire countries), and (iii) equally well for input and output measures.

Specifically, fields where advances are slower or where the resources involved are not too large (as pure mathematics) are much easier to quantify and manage than, for example, the life sciences where Federal investment is nowadays very large and for which the pace of change is staggeringly fast. Moreover, it is easier to forecast the impact of an increase in expenditure in the number of researchers, or the spending in new equipment, than it is to forecast what the impact will be in the number of publications, or the number of citations or even the shifting in the direction of the field.[1]

Our goal is to bring to bear on this problem concepts of statistical physics. Specifically, we guide our research by two concepts of fundamental importance in modern statistical physics: scale-invariance and universality. We try to identify robust, universal, characteristics of the evolution of S&T systems that can prove useful in forecasting the impact of changes in funding. Our proposed methods may be particularly useful for those fields of S&T (or for levels of aggregation) for which either not enough information is available or for which evolution is so fast that there isn't enough time to collect enough data to make an informed decision.

For our proposed approach to be useful, we first have to show that the principles of universality and scaling – which hold for complex physical systems – will also apply for S&T systems. In our preliminary work[13,14] we have tested these principles. We have quantified the production of research in a novel fashion inspired by our study of the growth dynamics of business firms. We studied the production of research from the point of view both of inputs (R&D funding) and of outputs (publications and patents). We also analyzed R&D systems of different countries to test the universality of our results.

Scaling and universality: Two concepts of modern statistical physics

Statistical physics deals with systems comprising a very large number of strongly-interacting subunits. Predicting the exact behavior of the individual subunit would be impossible, so one is limited to making statistical predictions regarding the collective behavior of the subunits. In the last century, statistical physics has begun to address systems (i) that are out of equilibrium, that is, systems driven by external "forces", and (ii) for which the exact interactions between the subunits comprising the system are not known. Recently, it has come to be appreciated that many such systems which consist of a large number of interacting particles obey universal laws that are independent of the microscopic details (will be given later in this section examples).

The finding, in physical systems, of universal properties that do not depend on the specific form of the interactions gives rise to the intriguing hypothesis that universal laws or results may also be present in economic and social systems. An often-expressed concern regarding the application of physics methods to the social sciences is that physical laws are said to apply to systems with a very large number of subunits (of order of $\approx 10^{23}$) while social systems comprise a much smaller number of elements. However, the "thermodynamic limit" is reached in practice for rather small systems. For example, in early computer simulations of gases or liquids reasonable results are already obtained for systems with 10^2-10^3 particles.

Background

First we introduce some of the advances that have occurred in our understanding of phase transitions and critical phenomena. Suppose we have a simple bar magnet. We know it is a ferromagnet because it is capable of picking up thumbtacks, the number of which is called the order parameter M. As we heat this system, M decreases and eventually, at a certain critical temperature T_c, it reaches zero: no more thumbtacks remain! In fact, the transition is remarkably sharp, since M approaches zero at T_c with infinite slope. Such singular behavior is an example of a "critical phenomenon."

The recent past of the field of critical phenomena has been characterized by several important conceptual advances, two of which are scaling and universality.

Scaling

The scaling hypothesis has two categories of predictions, both of which have been remarkably well verified by a wealth of experimental data on diverse systems. The first category is a set of relations, called *scaling laws*, that serve to relate the various critical-point exponents characterizing the behavior of functions such as M. For example, the magnetization M decays to zero at the critical temperature with an infinite slope which is quantified by a critical exponent. Similarly, other thermodynamic quantities display divergent behavior at the critical temperature. These behaviors are also quantified by different critical exponents. The scaling laws relate a sub-set of these critical exponents to the spatial dimension of the system under study.

The second category of predictions – which is more relevant to our study – is a sort of *data collapse*. To understand the importance of this fact, we will first take a short detour from our discussion, and present a simplified overview of some thermodynamic results.

Consider again the magnet bar we have been discussing. Remarkably, the macroscopic state of this magnet – that is, those characteristics of the magnet that can be measured without probing the state of individual atoms – can be uniquely characterized by just a few quantities: temperature T, applied magnetic field H, and magnetization M. Surprisingly, thermodynamics tells us that (any) one of these three quantities can be written as a function of the other two. This functional relationship – the so-called equation of state – can be written, for example, as $M=M(H,T)$. That is, given a temperature and applied magnetic field, one can calculate the magnetization, i.e., how many thumbtack our magnet picks up.

Close to the critical temperature, one can write the equation of state as $M=M(H,\tau)$, where $\tau\equiv(T-T_c)/T_c$ is a dimensionless measure of the deviation of the temperature T from the critical temperature T_c. Since $M(H,\tau)$ is a function of two variables, it can be represented graphically as M vs. τ for a sequence of different values of H.

The scaling hypothesis[18,19] predicts that all the curves of this family can be "collapsed" onto a single curve provided one plots not M vs. τ but rather a *scaled* M (M divided by H to some power) vs. a *scaled* τ (τ divided by H to some different power)

$$M(H,\tau) = H^a f\left(\frac{\tau}{H^b}\right). \tag{1}$$

The predictions of the scaling hypothesis are supported by a wide range of experimental work, and also by numerous calculations on model systems. Moreover, the general principles of scale invariance used here have proved useful in interpreting a number of other phenomena, ranging from elementary particle physics[15] and galaxy structure[16] to finance.[17]

Universality

The second theme goes by the name "universality". The utility of this concept is maybe better expressed through an analogy with the Mendeleev periodic table of atomic elements. Last century, Mendeleev notice that some elements shared similar physical and chemical properties. That observation prompted him to organize the known atomic elements at the time into a table in which atomic elements with similar properties occupied the same column. By organizing the elements into this table, Mendeleev found that some cells of this periodic table were left empty. Later it was found that those empty cells correspond to newly discovered atomic elements whose chemical and physical properties were well predicted by their position in the table.

Analogously, in statistical physics, it was found empirically that one could form a table in which the cells are occupied by a given system, e.g., a magnet, water at the critical point, or a polymer at its collapsing temperature. Surprisingly, these *a priori* rather different systems could be organized into a few classes, each class being described by the same scaling functions and the same set of scaling exponents.

This result is of great theoretical interest and it motivates an intriguing question: "Which features of this microscopic inter-particle force are important for determining critical-point exponents and scaling functions, and which are unimportant?"

Moreover, the discovery of universality in physical systems is also of great practical interest. Specifically, when studying a given problem, one may pick the most tractable system to study and the results one obtains will hold for all other systems in the same universality class.

Scale-invariance in systems outside of physics

At one time, it was imagined that the "scale-free" phenomena are relevant to only a fairly narrow slice of physical phenomena.[18] However, the range of systems that apparently display power law and hence scale-invariant correlations has increased dramatically in recent years, ranging from base pair correlations in non-coding DNA,[20] lung inflation,[21] plaque aggregation in Alzheimer's disease,[22] and interbeat intervals of the human heart[23] to complex systems involving large numbers of interacting subunits that display "free will," such as city growth,[24] stock price fluctuations[25] and currency exchange fluctuations.[26]

Moreover, and of greater relevance to the proposed work, we have recently shown that scaling and universality hold for economic organizations.[13,27–29] Namely, we found that the distributions of growth rates for both business firms and the gross domestic product (GDP) of entire countries are described by the same universal functional form and that the standard deviation of the distribution depends on organization size as a power law.

Scaling and universality in the growth of economic organizations

In the study of physical systems, the scaling properties of fluctuations in the output of a system often yield information regarding the underlying processes responsible for the observed macroscopic behavior.[18,19,30] With that in mind, we analyzed the fluctuations in the growth rates of different economic organizations.

Empirical results for business firms

In collaboration with an economist, Michael A. Salinger of Boston University, we investigated the growth dynamics of US business firms. A classic problem in industrial organizations is the size distribution of business firms.[31] For some time, it was assumed that firm size obeyed a rank-size law,[32] that is, that the distribution of sizes decays a power law of the size. In Figure 1a we show the distribution of log-sizes for US business firms, it is clear that the distribution has a fast decaying tail, inconsistent with a power law dependence.

We next consider the annual growth rate – that is to say, the fluctuation – of a firm's size,

$$g(t) \equiv \log\left(\frac{S(t+1)}{S(t)}\right), \qquad (2)$$

where $S(t)$ and $S(t+1)$ are the sales in US dollars of a given firm in the years t and $t+1$, respectively. We expect that the statistical properties of the growth rate g depend on S, since it is natural that the magnitude of the fluctuations g will decrease with S. Therefore, we partition the firms into bins according to their sales – the size of the firm. Figure 1b shows a log-linear plot of the probability distribution of growth rates for three sizes. In such a plot, a Gaussian distribution has a parabolic shape. It is apparent from the graph that the distributions are not Gaussian. Furthermore, it appears from the graph that the form of the distributions for the different sizes are similar. Indeed, Figure 1b suggests that the *conditional* probability density, $p(g|S)$, has the same functional form, with different widths, for all S.

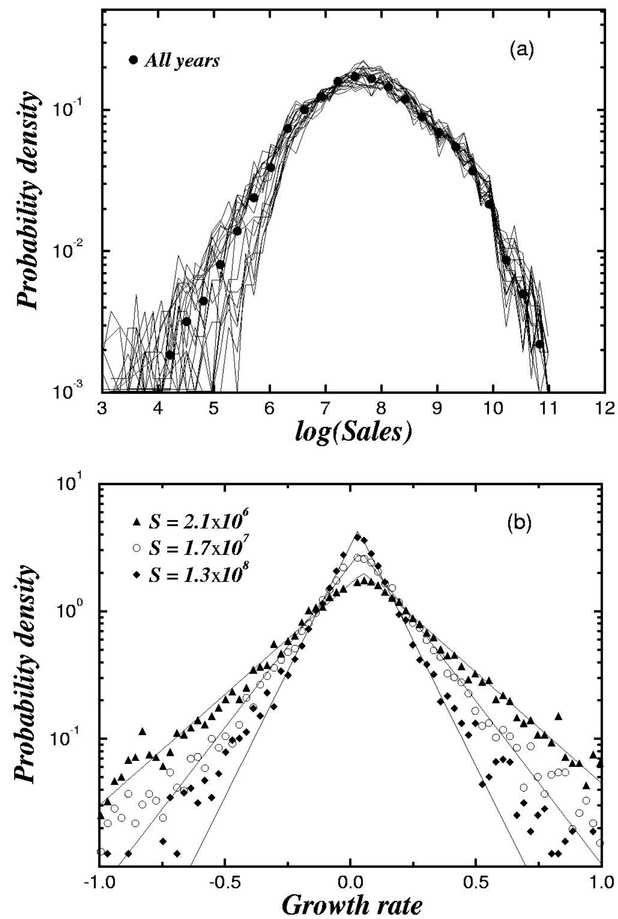


Figure 1. (a) Histogram of the sales S for publicly-traded manufacturing companies (with standard industrial classification index of 2000–3999) in the US for *each* of the years in the 1974–1993 period. All the values for sales were adjust to 1987 dollars by the GDP price deflator. Also shown (solid circles) is the average over the 20 years. It is visually apparent that the distribution is approximately stable over the period. (b) Probability density $p(r|S)$ of the growth rate r for all publicly-traded US manufacturing firms in the 1994 Compustat database with Standard Industrial Classification index of 2000–3999. The distribution represents all annual growth rates observed in the 19-year period 1974–1993. We show the data for three different bins of initial sales. The solid lines are exponential fits to the empirical data close to the peak. We can see that the wings are somewhat "fatter" than what is predicted by an exponential dependence.

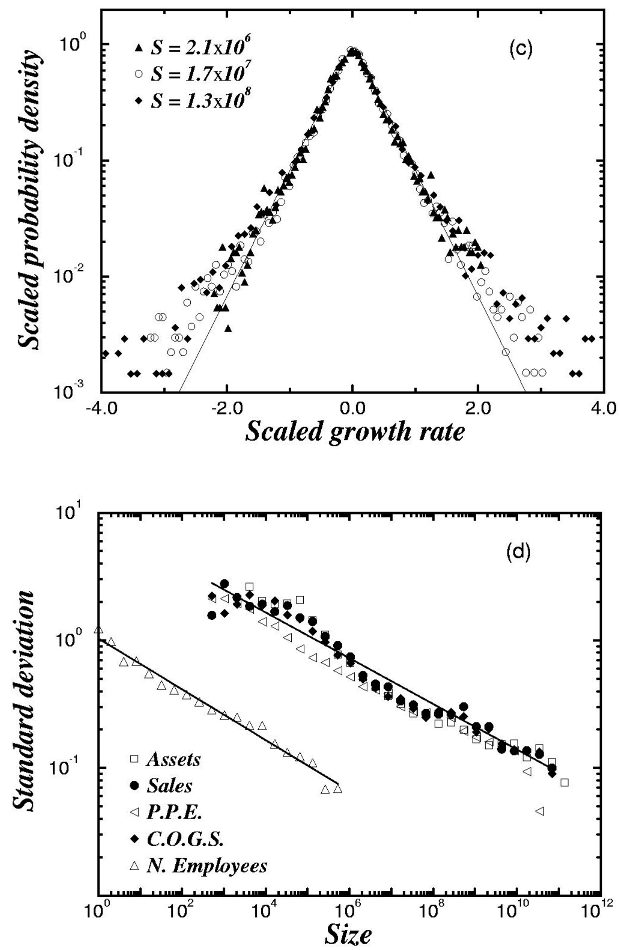


Figure 1. (c) Scaled probability density $p_{scat} \equiv \sigma p(g|S)$ as a function of the scaled growth rate $g_{scal} \equiv [g-\bar{g}]/\sigma$. The values were rescaled using the measured values of $\bar{g}$ and σ. All the data collapse upon the universal curve $p_{scal}=f(-|g_{scal}|)$. (d) Standard deviation of the 1-year growth rates σ for different definitions of the size of a company as a function of the initial values. We find that $\sigma \sim S^{-\beta}$. The straight lines are guides for the eye and have slopes 0.19.

To test if the conditional distribution of growth rates has a functional form independent of the size of the company, we plot the scaled quantities:

$$\sigma(S)p\left(\frac{g}{\sigma(S)}|S\right) vs. \frac{g}{\sigma(S)}, \tag{3}$$

where the function f is independent of S.

Figure 1c shows that the scaled conditional probability distributions "collapse" onto a single curve,[19] suggesting that $p(g|S)$ follows a universal scaling form [cf. Eq. (1)]

$$p(g|S) \sim \frac{1}{\sigma(S)} f\left(\frac{g}{\sigma(S)}\right). \tag{4}$$

We next calculate the standard deviation $\sigma(S)$ of the distribution of growth rates as a function of S. Figure 1d demonstrates that $\sigma(S)$ decays as a power law

$$\sigma(S) \sim S^{-\beta}, \tag{5}$$

with $\beta = 0.19 \pm 0.05$. One may ask if these results are only valid when the size of the firm is defined to be the sales. To test this possibility, we perform similar an analysis defining the size of the firms as (i) the number of employees, (ii) the assets, (iii) cost of goods sold (COGS), and (iv) plants, property and equipment (PPE). Figure 1d confirms that consistent results are obtained for all the above measures.

These results are intriguing for a number of reasons. First, we find consistent results for a set of firms belonging to a wide range of industries (from services in the bin for the smallest firms to oil and car companies in the bin for the largest firms). Second, we find consistent results for quite different types of measures of a firms' size, some such as COGS, PPE, assets and number of employees are input measures, while sales is an output measure. These two points suggest that universality is present in the growth dynamics of business firms. Third, we find power law scaling in the width of the distribution of growth rates, an unexpectedly "simple" results that suggests that simple mechanisms may explain our observations.

Empirical results for countries

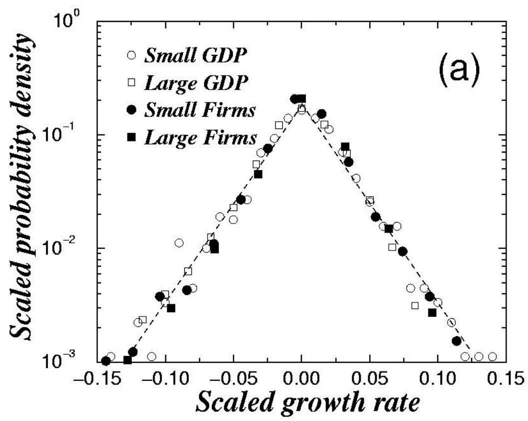
In collaboration with another economist, David Canning from The Queen's College in Dublin and Harvard University, we extended the analysis described in the previous subsections to the economy of countries. As earlier, we first consider the distribution of sizes S of a countries economy. Usually, the size of an economy is quantified by the gross domestic product (GDP) of the country.[33] Here, we detrend S by the world average growth rate, calculated for all the countries and years in our database.[34] We find that $p(\log S)$ is consistent with a Gaussian distribution, implying that $P(S)$ may be a log-normal. We also find that the distribution $P(S)$ does not depend on the time period studied.

Next, we calculate the distribution of annual growth rate g, as defined in Eq. (2), where $S(t)$ and $S(t+1)$ are the GDP of a country in the years t and $t+1$. As for business firms, we expect that the statistical properties of the growth rate g depend on S, since it is natural that the magnitude of the fluctuations g will decrease with S. Therefore, we partition the countries into bins according to their GDPs. We calculate the probability distribution of growth rates for three GDP sizes (small, medium and large) and find that the distributions are not Gaussian. Furthermore, as for business firms, the form of the distributions for the different sizes are consistent.

To test if the conditional distribution of growth rates has a functional form independent of the size of the company, we plot the scaled quantities (3). Figure 2a shows that the scaled conditional probability distributions "collapse" onto a single curve,[19] suggesting that $p(g|S)$ follows the universal functional form (4).

We next calculate the standard deviation $\sigma(S)$ of the distribution of growth rates as a function of S. Figure 2b demonstrates that $\sigma(S)$ decays as a power law, $\sigma(S) \sim S^{-\beta}$, with $\beta = 0.15 \pm 0.05$. We have also confirmed these results by a maximum-likelihood analysis.[35] In particular, we find that the log-likelihood of $p(g|S)$ being described by an exponential distribution – as opposed to a Gaussian distribution – is of the order of e^{600} to 1. Similarly, we test the log-likelihood of σ obeying Eq. (5). We find that Eq. (5) is e^{130} more likely than $\sigma(G) = \text{const}$, and that adding an additional nonlinear term to Eq. (5) does not increase the log-likelihood.

Surprisingly, we find that the same functional form appears to describe the probability distribution of annual growth rates for both the GDP of countries and the sales of firms; cf. Figure 2a. This result strongly suggests that universality, as defined in statistical physics, holds for the growth dynamics of economic organizations.



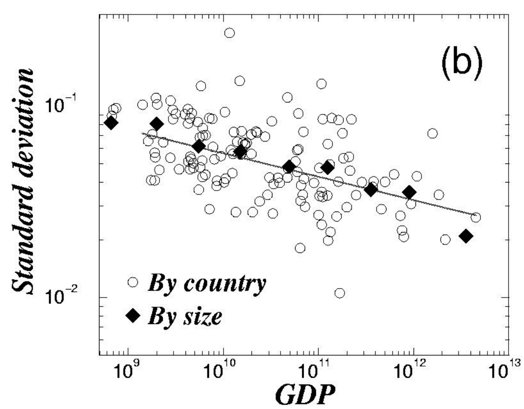


Figure 2. (a) Probability density function of annual growth rate for two subgroups with different ranges of G, where G denotes the GDP detrended by the average yearly growth rate. The entire database was divided into three groups: $6.9\times10^7 \le G < 2.4\times10^9$, $2.4\times10^9 \le G < 2.2\times10^{10}$, and $2.2\times10^{10} \le G < 7.6\times10^{11}$, and the figure shows the distributions for the smallest and largest groups. We consider only three subgroups in order to have enough events in each bin for the determination of the distribution. We plot the scaled probability density function, $\sigma(S)p(g/\sigma(S)|S)$, of the scaled annual growth rate, $(g-\bar{g})/\sigma(S)$ to show that all data collapse onto a single curve. (b) Standard deviation $\sigma(S)$ of the distribution of annual growth rates as a function of S, together with a power law fit (obtained by a least square linear fit to the logarithm of σ vs. the logarithm of S). The slope of the line gives the exponent β, with $\beta = 0.15$. We show the calculated standard deviation for two procedures: (i) for each individual country over the 42-yr period of the data, and (ii) for binned data according to size of GDP.

Modeling the growth dynamics of economic organizations

We next address the question of how to interpret our empirical results. We first note that an organization, such as a business firm, will comprise several subunits – the divisions of a firm. A reasonable zero-order approximation[36] is that the size of the different subunits comprising a firm will grow independently. Hence, we may view the growth of the size of each firm as the sum of the independent growth of subunits with different sizes. A model incorporating these assumptions[37] was recently proposed to describe the scale-invariant growth dynamics of different types of organizations.

Our model dynamically builds a diversified, multi-divisional structure, reproducing the fact that a typical firm passes through a series of changes in organization, growing from a single-product, single-plant firm, to a multi-divisional, multi-product firm.[38] The model reproduces a number of empirical observations for a wide range of values of parameters and provides a possible explanation for the robustness of the empirical results. Indeed, our model may offer a generic approach to explain power law distributions in other complex systems.

The model, illustrated in Figure 3, is defined as follows. A firm is created with a single division, which has a size $\xi_1(t=0)$. The size of a firm $S \equiv \Sigma_i \xi_i(t)$ at time t is the sum of the sizes of the divisions $\xi_i(t)$ comprising the firm. We define a minimum size $S_{\min}$ below which a firm would not be economically viable, due to the competition between firms; $S_{\min}$ is a characteristic of the industry in which the firm operates. We assume that the size of each division i of the firm evolves according to a random multiplicative process.[31] We define

$$\Delta\xi_i(t) \equiv \xi_i(t)\eta_i(t)\,, \tag{6}$$

where $\eta_i(t)$ is a Gaussian-distributed random variable with zero mean and standard deviation V independent of ξ_i. The divisions evolve as follows:

(i) If $\Delta\xi_i(t) < S_{\min}$, division i evolves by changing its size, and $\xi_i(t+1) = \xi_i(t) + \Delta\xi_i(t)$. If its size becomes smaller than $S_{\min}$ – i.e. if $\xi_i(t+1) < S_{\min}$ – then with probability p_a, division i is "absorbed" by division 1. Thus, the parameter p_a reflects the fact that when a division becomes very small it will no longer be viable due to the competition between firms.

(ii) If $\Delta\xi_i(t) > S_{\min}$, then with probability $(1-p_f)$, we set $\xi_i(t+1) = \xi_i(t) + \Delta\xi_i(t)$. With a probability p_f, division i does not change its size – so that $\xi_i(t+1) = \xi_i(t)$ – and an altogether new division j is created with size $\xi_j(t+1) = \Delta\xi_i(t)$. Thus, the parameter p_f reflects the tendency to diversify: the larger is p_f, the more likely it is that new divisions are created.

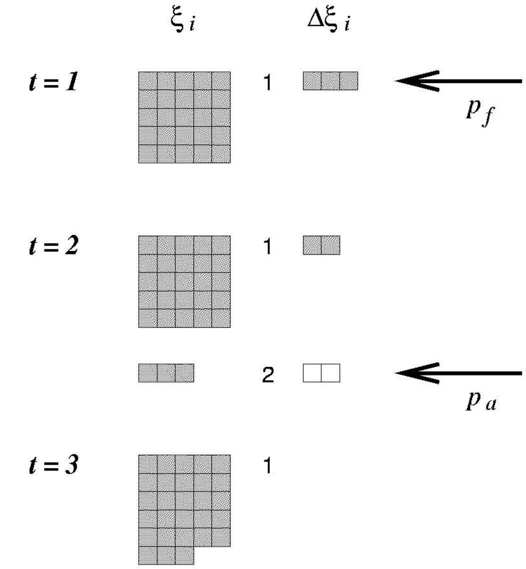


Figure 3. Schematic representation of the time evolution of the size and structure of a firm. We choose $S_{min} = 2$, and $p_f = p_a = 1.0$. The first column of full squares represents the size ξ_i of each division, and the second column represents the corresponding change in size $\Delta\xi_i$. Empty squares represent negative growth and full squares positive growth. We assume, for this example, that the firm has initially one division of size $\xi_1 = 25$, represented by a 5×5 square. At $t = 1$, division 1 grows by $\Delta\xi_1 = 3$. A new division, numbered 2, is created because $\Delta\xi_1 > S_{min} = 2$, and the size of division 1 remains unchanged, so for $t = 2$, the firm has 2 divisions with sizes $\xi_1 = 25$ and $\xi_2 = 3$. Next, divisions ξ_1 and ξ_2 grow by 2 and –2, respectively. Division 2 is absorbed by division 1, since otherwise its size would become $\xi_2 = 3-2 = 1$ which is smaller than S_{min}. Thus, at time $t = 3$, the firm has only one division with size $\xi_1 = 25+2+1 = 28$. Note that if division 1 would be absorbed, then division 2 would absorb division 1 and would then be renumbered 1. If, division 1 is absorbed and there are no more divisions left, the firm "dies."

The present model rests on a small number of assumptions. The three key assumptions are: (i) Firms tend to organize themselves into multiple divisions once they achieve a certain size. This assumption holds for many modern corporations.[38] (ii) There is a broad distribution of minimum scales in the economy. This assumption has also been verified empirically.[39] (iii) Growth rates of different divisions are independent of one another. For an economist, the third is the stronger of these assumptions. However, a recent study by John Sutton of the London School of Economics finds empirical support for this hypothesis.[36]

There are two features of our results that are perhaps surprising. First, although firms in our model consist of independent divisions, we do not find $\beta = 1/2$. To understand why $\beta<1/2$, suppose that the distribution of $s_m \equiv \ln S_{\min}$ is a Dirac δ-function. Although this assumption is unrealistic, it leads to an understanding of the underlying mechanisms in the model. In this case, it is a plausible assumption that the number of divisions will increase linearly with firm size, because the distribution of division sizes is narrow and confined between $S_{\min}$ and $S_{\min}/V$. This hypothesis is confirmed numerically, and we find (i) $\beta = 1/2$ and, (ii) that the distribution of the logarithm of firm sizes is still close to Gaussian, with a width w which is a function of the parameters of the model. Then, by integration of the distribution of the logarithm of firm sizes over s_m, we can estimate the value of β for the case of a broader distribution of s_m. Suppose that s_m follows some arbitrary distribution with width v. Averaging $\sigma^2(S)$ over this distribution, we find

$$\beta = \frac{w}{2(v+w)}. \tag{7}$$

To gain intuition on the results predicted by the expression, consider two representative cases: (1) $v=0$ implies $\beta = 1/2$, (2) $v=w$ implies $\beta = w/(4w) = 1/4$. For a wide range of the values of the model's parameters, we find $v>w$ implying that β is remarkably close to the empirical value $\beta \approx 0.2$.

Second, the distribution $p(g|S)$ is not Gaussian but "tent" shaped. We find this result arises from the integration of nearly-Gaussian distributions of the growth rates over the distribution of $S_{\min}$.

Predictions for an organizations' structure

We next address the question of the structure of a given firm. To this end, we calculate the probability density $\rho(\xi|S)$ to find a division of size ξ in a firm of size S. For the model, we find that the distribution ρ is scaled as a power law up to S^α and then it decays exponentially. Hence, we make the hypothesis that ρ obeys the scaling relation [cf. Eq. (1)]

$$\rho(\xi|S) \sim \frac{1}{S^\alpha} f\left(\frac{\xi}{S^\alpha}\right), \tag{8}$$

where $f(u) \sim u^\tau$ for $u<<1$ with $\tau \approx 2/3$.

The results described by Eq. (8) are in qualitative agreement with empirical studies[40] that show larger firms to be more diversified. Moreover, Eq. (8) implies that the number of independent subunits in a firm of size S scales as $S^{1-\alpha}$. Since N does not change much during a year and assuming that the subunits have similar sizes, we can apply the central limit theorem, from which it follows that $\sigma \sim N^{-1/2}$, leading to the testable scaling law

$$\beta = (1-\alpha)/2 . \tag{9}$$

Analysis of R&D growth at US universities

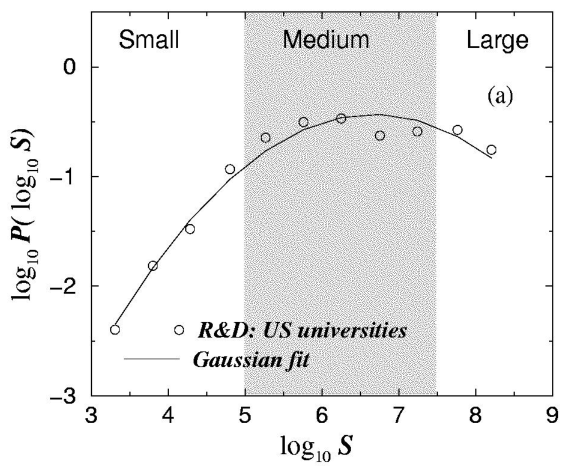
The impact of S&T activities on a country's economy led us to consider the question of the growth dynamic of R&D activities within a given national system. In particular, we hypothesized that the growth of R&D activities might mimic that of business firms. To test this hypothesis, we analyzed the fluctuations in the growth rates of university research activities, using five different measures of research activity. We studied the production of research both from the point of view of inputs – R&D funding – and outputs – publications and patents. First, we will consider an input quantity, the R&D expenditures of US research universities.[13]

Empirical analysis: R&D inputs

We analyzed an NSF database containing the annual R&D expenditures for science and engineering of more than 500 US universities[41] for the 17-year period 1979–1995 (≈12,000 data points). The expenditures – a measure of R&D inputs – are broken down by school and department. As in Eq. (2), the annual growth rate of R&D expenditures is defined as

$$g(t) \equiv \log\left(\frac{S(t+1)}{S(t)}\right), \tag{10}$$

where $S(t)$ and $S(t+1)$ are the R&D expenditures of a given university in the years t and $t+1$, respectively. Also, as before, we expect that the statistical properties of the growth rate g will depend on S. In fact, it is natural that the magnitude of the fluctuations in g will decrease with S. As shown in Figure 4a, we partition the universities into three classes according to the size of their R&D expenditures. Figure 4b shows a log-linear plot of the probability distribution of growth rates for the three size-classes of universities. In such a plot, a Gaussian distribution has a parabolic shape. It is apparent from the graph that the shapes are not parabolic, i.e., that the distributions are inconsistent with Gaussian statistics.



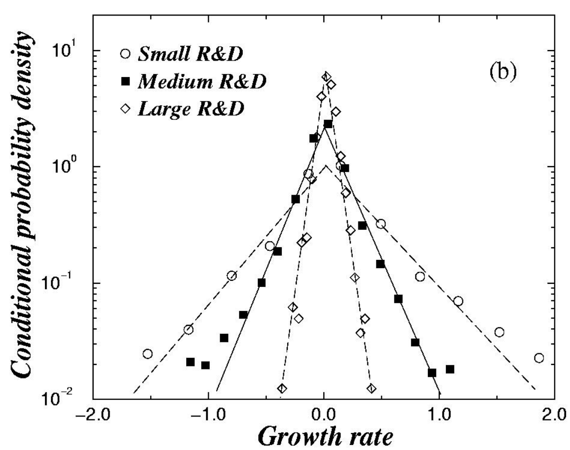


Figure 4. Growth dynamics of research activities at universities. (a) Histogram of the logarithm of the annual R&D expenditures of US universities for the 17-year period 1979–1995, expressed in 1992 US dollars. Here, S denotes the R&D expenditures detrended by inflation so that values for different years are comparable. The bins were chosen equally spaced on a logarithmic scale with bin size 0.5. The line is a Gaussian fit to the data, which is a prediction of Gibrat's theory.[31] (b) Conditional probability density function $p(g|S)$ of the annual growth rates g. For this plot the entire database is divided into three groups (depicted in (a) by different shades).

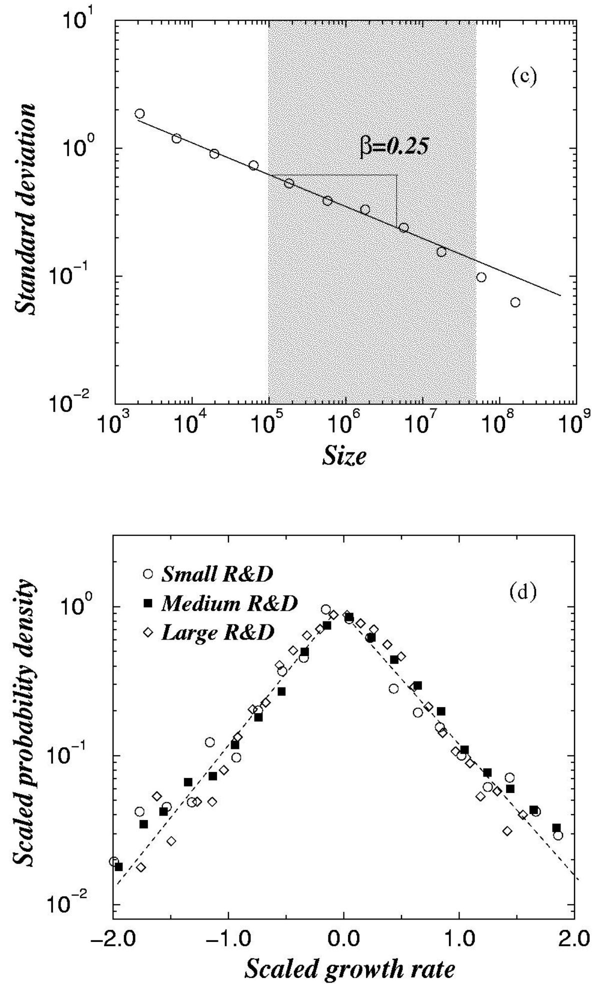


Figure 4. (c) Standard deviation $\sigma(S)$ of the distribution of annual growth rates as a function of S. The straight line is a power law fit to thedata, and its slope gives the exponent $\beta = 0.25 \pm 0.05$. (d) Scaled probability density function $p(g|S)/\sigma^{-1}(S)$ plotted against the scaled annual growth rate $(g-\bar{g})/\sigma(S)$ for the three groups defined in (b). Note that the scaled data collapse onto a single curve.

Based on the central limit theorem, one expects Gaussian distributions when one observes variables that are themselves the sum of a large number of nearly independent and nearly identically distributed random variables. While the growth rates of large universities are undoubtedly affected by a large number of variables many of which are largely independent of each other, we might expect only a small number of factors to dominate annual growth rates, especially for small universities.

Hence, it would appear that the distribution of growth rates should approach a Gaussian for the largest universities because the growth rate would be affected by more factors. In contrast to this expectation, it appears from the graph that the forms of the distributions for the different size classes are similar. Moreover, as we expected, the distribution is wider for smaller universities.

We next calculated the width $\sigma(S)$ of the distribution of growth rates as a function of S. Figure 4c shows a log-log plot of $\sigma(S)$ versus size. It is visually apparent that there is a strong dependence of the standard deviation of the growth rates on size. Moreover, it is striking that the relationship is nearly linear on a log-log plot, i.e., the width of the distribution decays as a power law with size $\sigma(S) \sim S^{-\beta}$. From the graph, we estimate that $\beta = 0.25 \pm 0.05$.

Figure 4b suggests that the *conditional* probability density, $p(g|S)$, has the same functional form, with different widths, for all S. To see whether they do, we produce what is called, in statistical physics, a "data collapse," shown in Figure 4d. We compute the distribution for different size classes and scale each by their standard deviation. As Figure 4d shows, the rescaled distributions "collapse" onto a single, "tent-shaped" curve.

Robustness of empirical results

R&D outputs: publications and patents. To test if these results for the growth of R&D expenditures are valid for other measures of research activity, we considered a measure of a university's research output: the number of papers published each year. We analyzed data from the *US University Science Indicators* published by ISI[42] for the 17-year period 1981–1997. This database records the number of papers published by the largest 112 US universities ($\approx$1,900 data points). We find that the analog of Figure 4 holds. Particularly striking is the fact that the same exponent value, $\beta \approx 1/4$, is found, and that the same functional form of $p(g|S)$ is displayed; cf. Figure 5.

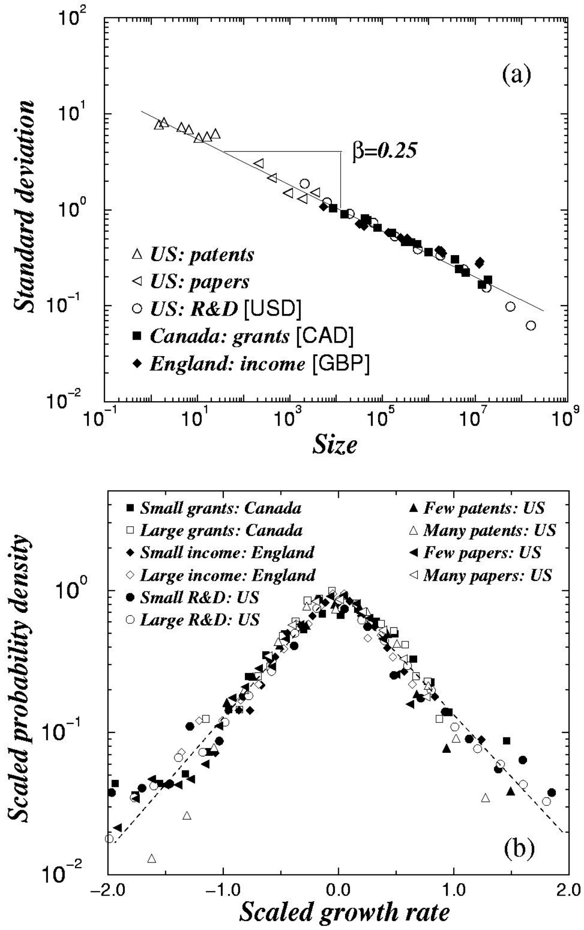


Figure 5. Robustness of empirical findings for the distribution of growth rates. (a) Standard deviation $\sigma(S)$ of the distribution of annual growth rates for different measures of research activities and different academic systems from the data in the five distinct databases analyzed: (i) the number of papers published each year at 112 US universities, (ii) the number of patents issued each year to 106 US universities, (iii) the R&D expenditures in US dollars of 719 US universities, (iv) the total amount in Canadian dollars of the grants to 60 Canadian universities, and (v) the external incomes in British pounds of 90 English universities. It is apparent that for all measures and all academic systems analyzed, we find a power law dependence – with the same exponent $\beta \approx 1/4$. The values of σ for the different measures were shifted vertically for better comparison of the estimates of the exponents. (b) The distribution of annual growth rates, scaled as in Figure 4d, for the five databases. We show the distribution of growth rates for 2 different groups, obtained in a way similar to that described in Figure 4b, for each of the five measures. The data appear to collapse onto a single curve, suggesting that the different measures have similar statistical properties.

Next, we analyzed another measure of R&D output: the number of patents issued to a university. We "manually" retrieve from the webpages of the *US Patent and Trademark Office's database*[43] the number of patents issued yearly to each of 106 universities for the 22-year period 1976–1997 (≈2,300 data points). We again confirm that the analog of Figure 4 holds, with the same exponent value, $\beta \approx 1/4$, and the same functional form of $p(g|S)$; cf. Figure 5.

Different academic systems. To test if our findings hold for different academic systems, we analyze two databases on research funding of English[44] and Canadian[45] universities. The same quantitative behavior is found for the distribution of growth rates and for the scaling of σ, with the same exponent value and the same functional form of $p(g|S)$; cf. Figure 5. Thus, our analysis of all five databases – which comprise a large number of universities from three different countries – indicates that the same quantitative results hold across different measures of research activity and academic systems.

Interpretation

A natural question that arises from our empirical findings for US universities is how to interpret their meaning. A first aspect we will focus on is the similarity of the results to those found for economic organizations. To understand this similarity we start with the observation that research is an expensive activity, and that the university must "offer" its research to external sources such as governmental agencies and business firms. Thus, an increase in R&D expenditures at university *A* and a decrease at university *B* implies that the funders of research increasingly choose their research from university *A* as opposed to university *B*.[46] This qualitative picture parallels the competition among different business firms, so it helps us to identify mechanisms that can lead to the same results for those apparently distinct types of organizations. Our results also suggest that peer review, together with government oversight, may lead to an outcome similar to that induced by market forces, where the analog of peer-review quality control may be consumer evaluation, and the analog of government oversight may be product regulation.

A second aspect is the origin of the similarity in the results for R&D expenditures and R&D outputs (patents and publications). Could it be that what we measure is merely a causal relationship rather than a universal physical law? It is plausible to assume some sort of causal relationship between expenditures and output, hence it is natural to wonder that the similarity of the results for input and output measures is just a reflection of that fact.* Nonetheless, it should be noted that the relationship between input and output is not a trivial one. If it was, then a lot of our problems would be solved and the solution to any problem would be "to throw money at it". There are reasons to believe that expenditures influence output and that output affects expenditures, at least, at the "microscopic" (research groups) or "mesoscopic" (universities) levels. The feedback between the two quantities are very likely a nonlinear process that we would want to understand at a basic level, since that may guide science policy in assessing responses to changes in funding or structure of S&T systems.

We feel it is particularly important to estimate the time delay between changes in expenditures and changes in output as a function of the scale of the S&T system. Only then can accurate measurements of the effect of changes in funding or priorities be correctly estimated.

Internal structure of US universities

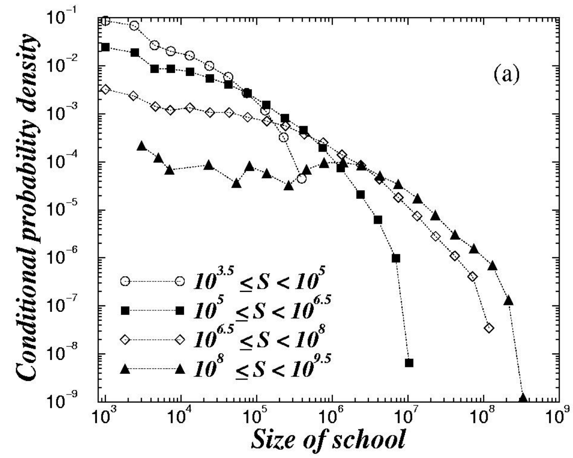
US universities are not monolithic entities but instead comprise different schools with varying degrees of autonomy. This fact prompts us to investigate the internal structure of US universities. As we showed previously, our model[37] for the growth of economic organizations makes several predictions regarding the size of the subunits comprising a given organization. Hence we can also test these predictions against the empirical data on US universities.

As discussed in connection with Eq. (8), we may quantify the internal structure of a university through the conditional probability density $\rho(\xi|S)$, that measures the probability to find a school of size ξ in a university of size S (Figure 6a). The model predicts that $\rho(\xi|S)$ obeys the scaling form[37]

$$\rho(\xi|S) \sim \frac{1}{S^{\alpha}} f\left(\frac{\xi}{S^{\alpha}}\right), \qquad (11)$$

where $f(u) \sim u^{-\tau}$ for $u \ll 1$, and $f(u)$ decays as a stretched exponential for $u \gg 1$.

* As pointed out by one of the Referees, the analysis of "publication propensity" of large European research institutions, reported in the 2nd Edition of the European Report on S&T Indicators, for instance, appears to suggest that there is neither a simple causal relationship between R&D expenditure and, say, publication output nor a unique "law" conditioning such relationships.



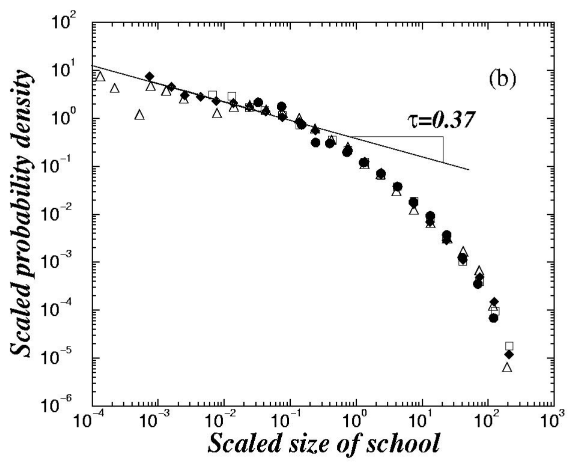


Figure 6. Statistical analysis of the units forming the internal structure of a university, the schools. (a) Conditional probability function $\rho(\xi|S)$ of finding a school of size ξ in a university of size S. To improve the statistics, we partition the universities by size into four groups. (b) To illustrate the scaling relation (11), we plot the scaled probability density $S^{\alpha}\,\rho(\xi/S^{\alpha}|S)$ versus the scaled size of the school ξ/S^{α}. In agreement with Eq. (11), we find that the scaled data fall onto a single curve.

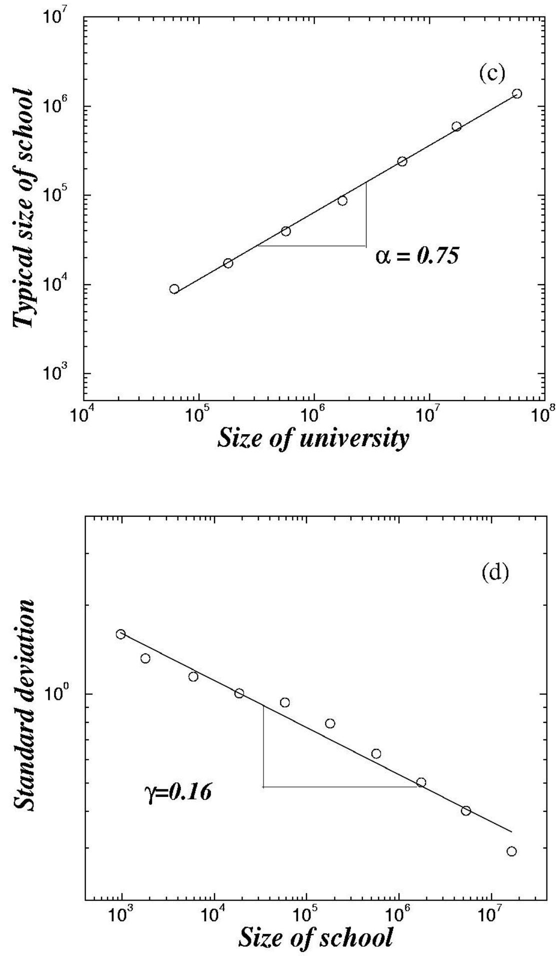


Figure 6. (c) Scaling of the typical size of a school in a university of a given size for different university sizes. The data obey a power law with exponent $\alpha = 0.75\pm0.05$. (d) Standard deviation w of the distribution of growth rates of schools versus school size ξ. The data obey a power law with exponent $\gamma = 0.16\pm0.05$. Using Eq. (15) and this value of γ, we obtain an independent estimate $\beta = 0.25\pm0.05$.

We find $\tau = 0.37 \pm 0.10$ (Figure 6b), and $\alpha = 0.75 \pm 0.05$ (Figure 6c). We test the scaling hypothesis (8) by plotting the scaled variables:

$$S^{\alpha} \rho\left(\frac{\xi}{S^{\alpha}} | S\right) vs. \frac{\xi}{S^{\alpha}}. \tag{12}$$

Figure 6b shows that all curves collapse onto a single curve, which yields the scaling function $f(u)$.

Equation (11) implies that the typical number of schools with research activities in a university of size S scales as $S^{1-\alpha}$, while the typical size of these schools scales as S^{α}. Hence, we can calculate how σ depends on S,

$$\sigma(S) \sim (S^{1-\alpha})^{-1/2} w(\xi). \tag{13}$$

In order to determine σ, we first find the dependence of w on ξ. Figure 6d shows that $w \sim \xi^{-\gamma}$ with $\gamma = 0.16 \pm 0.05$. Substituting into Eq. (13) and remembering that the typical size of the schools is S^{α}, we obtain

$$\sigma(S) \sim (S^{1-\alpha})^{-1/2} (S^{\alpha})^{-\gamma}, \tag{14}$$

which leads to the testable exponent relation

$$\beta = \frac{1-\alpha}{2} + \alpha\gamma. \tag{15}$$

For $\alpha \approx 3/4$ and $\gamma \approx 1/6$, Eq. (15) predicts $\beta \approx 1/4$, in surprising agreement with our empirical estimate of β from the five distinct databases analyzed (Figure 5a).

These results are intriguing for a number of reasons. First, they suggest that there are statistical laws describing the stationary structure of US universities. These laws impose constraints on the feasibility of achieving the intended goals of university administrators or governmental regulators. Second, our results suggest that for very large universities the largest school has a considerable smaller size than the size of the university while for small universities a single school may dominate. Third, as expected we find that even the schools cannot be modelled as monolithic structures and that it is important to study R&D at the level of the departments and of the research groups. Finally, one may hypothesize that scaling laws may hold for the growth of schools, departments and research groups.

Conclusion

In this paper, we have described a style of research that may be quite new in the field of science policy. We believe that this new approach may shed new light on the behavior and characteristics of S&T systems, answering questions such as "To what extent do S&T systems at different scales (countries, universities, different financing modes) obey the same underlying evolutionary laws?"

Understanding these processes and the data characterizing them is of great relevancy not only in S&T studies but also for science policy. Indeed, OECD countries are increasingly stressing performance because research funding is becoming more and more an instrument in safeguarding long term economic competitiveness.

*

We thank M. *Barthélemy*, S. V. *Buldyrev*, D. *Canning*, X. *Gabaix*, S. *Havlin*, P. Ch. *Ivanov*, H. *Kallabis*, Y. *Lee*, H. *Leschhorn*, F. *Liljeros*, P. *Maass*, M. *Meyer*, B. *Roehner*, M. A. *Salinger*, M. H. R. *Stanley* and especially Henk *Moed* and Marc *Luwel* for stimulating discussions. We also thank N. *Bayers*, E. *Garfield*, and especially R. E. *Hudson* for help with obtaining the ISI database. The CPS is supported by NSF and NIH.

References

1. National Science Board, *Science & Engineering Indicators - 2000,* National Science Foundation, Arlington VA, 2000.
2. E. MANSFIELD, Academic research and industrial-innovation, *Research Policy*, 20 (1991) 1–12.
3. A. PAKES, K. L. SOKOLOFF, Science, technology, and economic growth, *Proc. Nat. Ac. Sci. USA*, 93 (1996) 12655–12657.
4. H. GRUPP, The links between competitiveness, firms' innovative activities and public R&D support in Germany: An empirical analysis, *Technol. Anal. Strateg.*, 9 (1997) 19–33.
5. Program of *Sixth International Conference on Science and Technology Indicators* (S&T 2000), CWTS, Leiden, 2000.
6. E. GARFIELD, *Citation Indexing: Its Theory and Applications in Science, Technology and Humanities,* Wiley, New York, 1979.
   F. NARIN, Evaluative Bibliometrics. *The Use of Publication and Citation Data in the Evaluation of scientific activity*, National Science Foundation, Washington DC, 1976.
   A. SCHUBERT, W. GLÄNZEL, T. BRAUN, Scientometric datafiles: A comprehensive set of indicators on 2649 journals and 96 countries in all major science fields and subfields 1981–1985, *Scientometrics*, 16 (1989) 3–478.
   M. B. ALBERT, D. AVERY, F. NARIN, P. MCALLISTER, Direct validation of citation counts as indicators of industrially important patents, *Research Policy*, 20 (1991) 251–259.
   H. GRUPP, The measurement of technical performance of innovations by technometrics and its impact on established technology indicators, *Research Policy*, 23 (1994) 175–193.
   P. O. SEGLEN, Citation rates and journal impact factors are not suitable for evaluation of research, *Acta. Orthop. Scand.*, 69 (1998) 224–229.

S. TEITEL, Scientific publications, research-and-development expenditures, country size, and per-capita income: A cross-section analysis, *Technol. Forecast. Soc.*, 46 (1994) 175–187.
A. F. J. VAN RAAN, Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises, *Scientometrics*, 36 (1996) 397–420.
F. NARIN, D. OLIVASTRO, Linkage between patents and papers: An interim EPO/US comparison, *Scientometrics*, 41 (1998) 51–59.
A. F. J. VAN RAAN, Scientometrics: State-of-the-art, *Scientometrics*, 38 (1997) 205–218.
E. C. M. NOYONS, M. LUWEL, H. F. MOED, Combining mapping and citation analysis for evaluative bibliometric purposes, *J. Am. Soc. Inf. Sci.*, 50 (1999) 115–131.
7. K. BUCHHOLZ, Criteria for the analysis of scientific quality, *Scientometrics*, 32 (1995) 195–218.
A. T. BALABAN, How should citations to articles in high- and low-impact journals be evaluated, or what is a citation worth? *Scientometrics*, 37 (1996) 495–498.
H. F. MOED, T. N. VAN LEEUWEN, J. REEDIJK, A critical analysis of the journal impact factors of Angewandte Chemie and the Journal of the American Chemical Society – Inaccuracies in published impact factors based on overall citations only, *Scientometrics*, 37 (1996) 105–116.
P. VINKLER, Model for quantitative selection of relative scientometric impact indicators, *Scientometrics*, 36 (1996) 223–236.
W. GLÄNZEL, The need for standards in bibliometric research and technology, *Scientometrics*, 35 (1996) 167–176.
M. DE MARCHI, M. ROCCHI, Summing up approaches to the study of science and technology indicators, *Scientometrics*, 46 (1999) 39–49.
8. H. F. MOED, 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. R. J. W. TIJSSEN, Th. N. VAN LEEUWEN, B. VERSPAGEN, H. HOLLANDERS, *1998 Science and Technology Indicators: Summary*, A joint publication for the Dutch Ministry of Education, Culture and Science, Zoetermeer, 1998.
10. M. LUWEL, E. C. M. NOYONS, H. F. MOED, Bibliometric assessment of research performance in Flanders: Policy background and implications, *R&D Management*, 29 (1999) 133–141.
11. H. F. MOED, Th. N. VAN LEEUWEN, M. S. VISSER, Trends in publication output and impact of universities in the Netherlands, in: *Proceedings of the 5th International Conference on Science and Technology Indicators*, edited by H. F. MOED, Hinxton, Cambridge UK, 1998, *Research Evaluation*, 8 (1999) 60–67.
12. M. LUWEL, A bibliometric profile of Flemish research in natural, life and technical sciences, *Scientometrics*, 47 (2000) 281–302.
13. V. PLEROU, 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.
14. H. F. MOED, M. LUWEL, Science policy: the business of research, *Nature*, 400 (1999) 411–412.
15. R. JACKIW, Introducing scale symmetry, *Phys. Today*, 25(1) (1972) 23–27.
16. P. J. E. PEEBLES, *The Large-Scale Structure of the Universe*, Princeton University Press, Princeton, 1980.
17. R. N. MANTEGNA, H. E. STANLEY, Scaling behavior in the dynamics of an economic index, *Nature*, 376 (1995) 46–49.
R. N. MANTEGNA, H. E. STANLEY, Turbulence and exchange markets, *Nature*, 383 (1996) 587–588.
18. H. E. STANLEY, *Introduction to Phase Transitions and Critical Phenomena*, Oxford, Oxford University Press, 1971.

19. H. E. STANLEY, Scaling, universality, and renormalization: Three pillars of modern critical phenomena, *Rev. Mod. Phys.*, 71 (1999) S358–S366 [Special Issue for the Centennial of the American Physical Society].
20. C.-K. PENG, S. V. BULDYREV, A. L. GOLDBERGER, S. HAVLIN, F. SCIORTINO, M. SIMONS, H. E. STANLEY, Long-range correlations in nucleotide sequences, *Nature*, 356 (1992) 168–171.
A. ARNEODO, E. BACRY, P. V. GRAVES, J. F. MUGY, Characterizing long-range correlations in DNA-sequences from wavelet analysis, *Phys. Rev. Lett.*, 74 (1995) 3293–3296.
21. B. SUKI, A.-L. BARABÁSI, Z. HANTOS, F. PETÁK, H. E. STANLEY, Avalanches and power law behaviour in lung inflation, *Nature*, 368 (1994) 615–618.
22. B. T. HYMAN et al., quantitative analysis of senile plaques in Alzheimer disease: Observation of log-normal size distribution and of differences associated with apolipoprotein e genotype and trisomy 21 (Down Syndrome), *Proc. Natl. Acad. Sci. USA*, 92 (1995) 3586–3590.
L. CRUZ et al., Aggregation and disaggregation of senile plaques in alzheimer disease, *Proc. Natl. Acad. Sci. USA*, 94 (1997) 7612–7616.
R. B. KNOWLES et al., Plaque-induced neural network disruption in Alzheimer's disease, *Proc. Natl. Acad. Sci. USA*, 96 (1999) 5274–5279.
23. C.-K. PENG, S. HAVLIN, H. E. STANLEY, A. L. GOLDBERGER, Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series, *Chaos*, 5 (1995) 82–87.
L. A. N. AMARAL, A. L. GOLDBERGER, P. Ch. IVANOV, H. E. STANLEY, Scale-independent measures and pathologic cardiac dynamics, *Phys. Rev. Lett.*, 81 (1998) 2388–2391.
P. Ch. IVANOV, L. A. N. AMARAL, A. L. GOLDBERGER, H. E. STANLEY, Stochastic feedback and the regulation of biological rhythms, *Europhys. Lett.*, 43 (1998) 363–369.
P. Ch. IVANOV, L. A. N. AMARAL, A. L. GOLDBERGER, S. HAVLIN, M. G. ROSENBLUM, Z. STRUZIK, H. E. STANLEY, Multifractality in human heartbeat dynamics, *Nature*, 399 (1999) 461–465.
24. M. BATTY, P. LONGLEY, *Fractal Cities*, San Diego: Academic Press, 1994.
H. MAKSE, S. HAVLIN, H. E. STANLEY, Modeling urban growth patterns, *Nature*, 377 (1995) 608–612.
X. GABAIX, Zipf's law for cities: An explanation, *Quarterly J. Econ.*, 114 (1999) 739–767.
25. B. B. MANDELBROT, The variation of certain speculative prices, *J. Business*, 36 (1963) 394–419.
A. PAGAN, The econometrics of financial markets, *J. Empirical Finance*, 3 (1996) 15–102.
T. LUX, The stable paretian hypothesis and the frequency of large returns: An examination of major German stocks, *Applied Financial Economics*, 6 (1996) 463–475.
U. A. MULLER, M. M. DACOROGNA, O. V. PICTET, Heavy tails in high-frequency financial data, in: *A Practical Guide to Heavy Tails*, edited by R. J. ADLER, R. E. FELDMAN, M. S. TAQQU, Birkhäuser Publishers, 1998, pp. 83–311.
V. PLEROU, P. GOPIKRISHNAN, L. A. N. AMARAL, M. MEYER, H. E. STANLEY, Scaling of the distribution of price fluctuations of individual companies, *Phys. Rev. E*, 60 (1999) 6519–6529.
P. GOPIKRISHNAN, V. PLEROU, L. A. N. AMARAL, M. MEYER, H. E. STANLEY, Scaling of the distributions of fluctuations of financial market indices, *Phys. Rev. E*, 60 (1999) 5305–5316.
26. U. A. MULLER, M. M. DACOROGNA, R. B. OLSEN, O. V. PICTET, M. SCHWARZ, C. MORGENEGG, *Journal of Banking and Finance*, 14 (1990) 1189–1195.
27. M. H. R. STANLEY, L. A. N. AMARAL, S. V. BULDYREV, S. HAVLIN, H. LESCHHORN, P. MAASS, M. A. SALINGER, H. E. STANLEY, Scaling behaviour in the growth of companies, *Nature*, 379 (1996) 804–806.
28. L. A. N. AMARAL, S. V. BULDYREV, S. HAVLIN, H. LESCHHORN, P. MAASS, M. A. SALINGER, H. E. STANLEY, M. H. R. STANLEY, Scaling behavior in economics: I. Empirical results for company growth, *J. Phys. I France*, 7 (1997) 621–633.
29. Y. LEE, L. A. N. AMARAL, D. CANNING, M. MEYER, H. E. STANLEY, Universal features in the growth dynamics of complex organizations, *Phys. Lev. Lett.*, 81 (1998) 3275–3278.
30. H. E. STANLEY, Power laws and universality, *Nature*, 378 (1995) 554–555.

31. R. GIBRAT, *Les Inégalités Economiques*, Sirey, Paris, 1931.
Y. IJIRI, H. A. SIMON, *Skew Distributions and the Sizes of Business Firms*, North Holland, Amsterdam, 1977.
B. H. HALL, The Relationship between firm size and firm growth in the U.S. manufacturing sector, *The J. Indust. Econ.*, 35 (1987) 583–606.
J. SUTTON, Gibrat's legacy, *J. Econ. Lit.*, 35 (1997) 40–59.
32. V. PARETO, *Cours d'Economie Politique*, Lausanne and Paris, 1897.
P. LÉVY, *Théorie de l'Addition des Variables Aléatoires*, Gauthier-Villars, Paris, 1937.
G. K. ZIPF, *Human Behavior and the Principle of Least Effort*, Addison-Wesley, Cambridge MA, 1949.
G. R. CARROLL, National city-size distribution: What do we know after 67 years of research? *Progress in Human Geography*, VI (1982) 1–43.
33. S. N. DURLAUF, On the convergence and divergence of growth rates, *The Economic J.*, 106 (1996) 1016–1018.
34. R. SUMMERS, A. HESTON, The Penn World Tables (Mark 5): An expanded set of international comparisons, 1950–1988, *Quarterly J. Economics*, 106 (1991) 327–368.
35. W. H. PRESS et al., *Numerical Recipes*, 2nd ed., Cambridge University Press, Cambridge, 1992.
36. J. SUTTON, *The Variance of Firm Growth Rates: The 'Scaling' Puzzle,* Working paper, London School of Economics, 2000.
37. L. A. N. AMARAL, 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. Letters*, 80 (1998) 1385–1388.
38. A. CHANDLER, *Strategy and Structure*, MIT Press, Cambridge, 1962.
39. M. GORT, *Diversification and Integration in American Industry*, Princeton University Press, Princeton, 1962.
40. B. JOVANOVIC, The diversification of production, *Brookings Pap. Econ. Ac.: Microeconomics (1)*, VI (1993) 197–247.
41. NATIONAL SCIENCE FOUNDATION, DIVISION OF SCIENCE RESOURCES STUDIES, *Academic Research and Development Expenditures*, NSF, Arlington VA, 1998.
42. *United States University Science Indicators on Diskette, 1981–1997*, Institute for Scientific Information, Philadelphia, 1998.
43. *United States Patent and Trademarks Office Databases 1976–1997* [http://www.uspto.gov].
44. HIGHER EDUCATION FUNDING COUNCIL FOR ENGLAND, *The 1996 Research Assessment Exercise*, HEFCE, Bristol, 1996.
45. NATURAL SCIENCES AND ENGINEERING RESEARCH COUNCIL OF CANADA, *NSERC Grant Database for 1991–1998*, NSERC, Ottawa, 1999.
46. H. F. MOED, M. LUWEL, J. A. HOUBEN, H. VAN DEN BERGHE, E. SPRUYT, Funding and research performance, *Nature*, 392 (1997) 119–120.

Received February 16, 2001.

Address for correspondence:
L. A. N. AMARAL
Center for Polymer Studies and Department of Physics,
Boston University, Boston, MA 02215, USA
E-mail: amaral@buphy.bu.edu