Scale-free behavior and universality in random fragmentation and aggregation

Jayanth R. Banavar,[1] Paolo De Los Rios,[2] Alessandro Flammini,[3,4] Neal S. Holter,[1] and Amos Maritan[4,5]
[1]*Department of Physics, The Pennsylvania State University, University Park, Pennsylvania 16802, USA*
[2]*Laboratoire de Biophysique Statistique, ITP-SB, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland*
[3]*International School for Advanced Studies, Trieste, Italy*
[4]*Unità INFM, Trieste and Abdus Salam International Center for Theoretical Physics, Trieste, Italy*
[5]*INFM and Dipartimento di Fisica "G. Galilei," Universita di Padova, Via Marzolo 8, 35131 Padova, Italy*
(Received 1 October 2003; published 31 March 2004)

Two distinct mechanisms underlying the existence of power-law distributions are presented: the distribution is stationary under the process of merging and splitting of classes and the distribution of the entities under study is invariant under changes of the classification scheme. We provide an explanation for the ubiquitous inverse n relationship in the species abundance relationship in ecology and the $1/n^2$ distribution of company sizes based on the minimum impact principle.

DOI: 10.1103/PhysRevE.69.036123 PACS number(s): 89.75.Da, 64.60.Ak, 89.65.Ef, 87.23.Cc

Scale-free power-law distributions are observed commonly [1–8] and there have been many efforts to understand their ubiquity. The primary focus of our paper is to elucidate two mechanisms for understanding the origins of power-law behavior. The first arises from a generalization of the Smoluchowski equation [9], which has been previously studied in the context of random fragmentation and agglomeration [10–16] and in the formation of groups of animals [17]. Such studies of rupture and aggregation have applications in condensed matter physics, statistical physics, and cosmology and it is known that they can produce power-law distributions. We seek to describe mergers and spin-offs in, say, a world of companies, and show numerically and analytically that one obtains, in steady state, a stable, nontrivial, fixed probability distribution of company sizes with nonuniversal power-law behavior.

A system at a critical point [3] is characterized by scale invariance leading to power-law behavior and looks much the same independent of the resolution with which you view it. Such a system is well defined and there is no subjectivity in measuring its attributes. In contrast, we show that, in situations in which there is some subjectivity in the categorization, the very fact that a consistent pattern is observed in a robust manner imposes a strong constraint on the nature of the distribution and implies that the underlying pattern exhibits algebraic behavior. We call this second mechanism the principle of recategorization invariance, which we illustrate with the species abundance relationship in ecology.

Finally, we present a scaling analysis [3], which together with a minimum impact principle is shown to lead to distinct universal [3] power-law relationships, without any fine-tuning of parameters [2], in the company size distributions and in ecology in excellent accord with data [18–21].

We have performed numerical simulations to determine the distributions of company sizes [18] after a series of splits and mergers and, in all cases in which stationarity is attained, a power-law distribution of company sizes is obtained, but with exponents whose values depend on the rules employed. The number of companies, $s(n)\,dn$, with number of employees between n and $n+dn$ is found to scale as $n^{-\theta}$, with the power-law behavior being cut off for sufficiently large values of n.

In a split, a randomly chosen company is divided into two, preserving the total number of people. In a merger, all of the people from two randomly chosen different companies are combined into a single company. Therefore, the probability that two companies of sizes s_1 and s_2 participate in a merging event is proportional to the product of the number of companies of sizes s_1 and s_2. Iterations consisting of a single split, with probability p, or a single merger, with probability $1-p$, are performed until the distribution of companies reaches a statistical steady state. In principle, one could consider the general situation in which the probability of splitting and merging depends on the sizes of the two parts involved in the event. The corresponding equations can be written down, but it is hard to obtain an analytical solution. We restrict our analysis below to some special cases for which explicit results can be obtained. The common features of the solutions we find are suggestive of their generality.

We worked with two different types of splits. The first, random split, consists of selecting a company at random and splitting it into two companies, one with a size randomly selected between zero and the size of the initial company, and the other with the remainder of the people. The second split type, Equal split, consists of selecting a company at random and splitting it into two of exactly half the size of the original. In both cases, there is no minimum company size and the size is not required to be an integer. Similarly, we have also considered two types of mergers. First, random merger, consists of selecting two companies at random and replacing them with a single company with a size equal to the sum of the two original companies. The second type, closest merger, consists of selecting one company at random and then merging it with the company closest to it in size. We choose a maximum company size M so that mergers which lead to companies with more than M individuals are not permitted. This could model the action of an antitrust authority. Note that the overall population is finite (the total number of workers in the case of companies, for example) so that mergers leading to companies larger than the overall population cannot take place.

The master equation for the random-splitting–random-merging process is

$$s^{i+1}(n)-s^i(n)=p\left(-s^i(n)+2\int_n^\infty \frac{s^i(m)}{m}dm\right)+(1-p)\times\left(\int_0^n s^i(m)s^i(n-m)dm\right)-(1-p)\times\left(2s^i(n)\int_0^{M-n}s^i(m)dm\right). \quad (1)$$

The first two terms on the right-hand side represent the loss and the gain due to splitting while the third and fourth terms describe the gain and the loss due to merging.

On postulating a power-law form for $s(n)$ in the stationary situation and equating terms in leading order, one obtains the exponent

$$\theta=\frac{2p}{2-p}. \quad (2)$$

Note that when $p>2/3$, $\theta>1$ and there is no stationary solution. Similarly, for the equal-splitting–random-merging case, the term $2\int_n^\infty[s^i(m)/m]dm$ in Eq. (1) is replaced by $4s^i(2n)$, leading to

$$\theta=2+\ln\left(\frac{p}{2-p}\right)/\ln(2). \quad (3)$$

Finally, for the random-splitting–closest-merge case, the master equation is

$$s^{i+1}(n)-s^i(n)=p\left(-s^i(n)+2\int_n^\infty \frac{s^i(m)}{m}dm\right)+(1-p)\times[\tfrac{1}{2}s^i(n/2)-2s^i(n)], \quad (4)$$

leading to the p-θ relation:

$$p=\frac{2-2^{\theta-1}}{1+2/\theta-2^{\theta-1}}. \quad (5)$$

We have verified these predictions with computer simulations. Some representative cases are shown in Fig. 1.

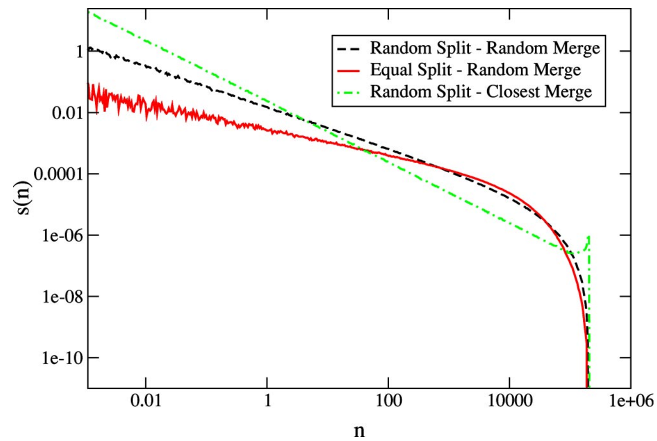

FIG. 1. Plot of $s(n)$ vs company size n for three distinct split-merger simulations; $s(n)dn$ is the fraction of companies of size (number of employees) between n and $n+dn$. All three cases were run with 20 companies and a total of 200 000 employees. The value of p in each case is 1/2 and the observed exponents (0.66, 0.42, and 1.0) are in good accord with the analytic predictions of 2/3, 0.415, and 1 for the random-split–random-merge, equal-split–random-merge, and the random-split–closest-merge cases, respectively.

The framework of the Smoluchowski equation [9] is also relevant for understanding the principle of recategorization invariance. In order to study distributions of quantities of interest, one needs to categorize or bin them. Often, this can be done in an objective manner but there are situations in which this is not possible. For example, if one wished to study the distribution of papers published in scientific journals in various subject categories or the distribution of people in various employment sectors, different analysts of the same data could choose distinct categorization schemes and could presumably get different distributions. If the distributions are indeed qualitatively different, then their behavior is not robust and depends on the specific categorization scheme. More interestingly, if the distributions are in the same class or equivalently are described by the same family of mathematical functions independent of the specific scheme employed, it would be natural to think of this as a pattern worth understanding.

Let us illustrate recategorization invariance with a fragmentation example relevant to ecology. How might one categorize organisms into species? There is clearly no unique answer to this question [22–24] except that, within any categorization, organisms belonging to a given species ought to be closer to each other than to organisms in other species. In order to simplify the situation and derive the consequences of this flexibility in classification, we envision starting with a coarse definition of the species and ask how the distribution evolves on recategorizing the organisms into species with a finer distinction between them. In other words, we consider carrying out a simple splitting procedure [25] in which, for example, each species is divided into two and the integer population in the original species (measured in convenient units so that, for example, the unit of population corresponds to extinction threshold) is divided randomly, for simplicity, into positive integer populations of two new species.

We do not concern ourselves with species having a population of just 1—such a species cannot, of course, be split any further and may be thought of as one that goes extinct. A robust law or a consistent pattern is one that ought to be observed independently of the precise definition of the species. When the total number of species is large, our simulations show that the splitting procedure is well described by the mean field recursion relation for $s^{i+1}(n)$, the number of species with a population of n at the $(i+1)$th iteration:

$$s^{i+1}(n)=2\sum_{m>n}\frac{s^i(m)}{m-1}. \quad (6)$$

The stable fixed probability distribution of this recursion is one in which each organism belongs to a distinct species. Any initial distribution will eventually reach this stable distribution after infinite iterations of the splitting process. Another fixed probability solution (albeit unstable) to Eq. (6) is a power-law distribution. This follows readily from considering an approximate integral representation of Eq. (6) written as $s^{i+1}(n)=2\int_n^\infty [s^i(m)/m]dm$ and noting that $s^i(m)\sim m^{-\theta}$ leads to $s^{i+1}(n)\sim n^{-\theta}$.

Figure 2 shows the results of integrating the recursion relation for three commonly studied classes of distributions [19,20], the canonical log-normal, the broken stick, and a power-law form. The power-law distribution retains the most fidelity to its functional form on successive iterations. The canonical log-normal form is somewhat robust under iteration with just a weak variation in the adjustable parameter. Note, however, that a power-law is a reasonable approximation to the tail of a canonical log-normal distribution.

It is important to emphasize that the species abundance relationship arises from evolution and natural selection, and the observed regularities do not derive from the way in which one categorizes the species. Instead, the distributions which have been commonly put forward to characterize the relative abundance of species (the log-normal and the power-law) are invariant under recategorization of the quantities being studied and are therefore consistent with the robust observability requirement discussed here.

Our studies of successive splitting and mergers and the principle of invariance under recategorization both lead to power-law distributions but with no unique exponent. We now turn to a scaling analysis [3] of such distributions. Let N represent the total number of individuals across all species:

$$N=\int_1^\infty s(n)n\,dn. \tag{7}$$

Our simulations have shown that the number of companies (species) of size n (with n individuals), $s(n;N)$, is a homogeneous function of the type

$$s(n;N)\sim n^{-\theta}F(n/n_0(N)). \tag{8}$$

The scaling function $F(x)$ approaches a constant value for x small compared to 1 and becomes zero when x is large compared to 1, so n_0 is the upper limit cutoff of the algebraic behavior of $s(n;N)$. A sample of the results of our simulations is shown in the collapse plot of $n^\theta s(n;N)$ versus $n/n_0(N)$ in Fig. 3. This scaling form is well captured by the Fisher log series [21] which states that the number of species with a population n is proportional to $e^{-n/n_0}/n$.

The total number of individuals is given by

$$N=\int_1^\infty s(n)n\,dn\sim n_0^{2-\theta}\int_{1/n_0}^\infty x^{1-\theta}F(x)dx. \tag{9}$$

In order for n_0 to diverge as $N\rightarrow\infty$ and the scaling regime to be extended indefinitely, $\theta\leqslant 2$. The total number of companies (species) is equal to

$$C=\int_1^\infty s(n)dn\sim n_0^{1-\theta}\int_{1/n_0}^\infty x^{-\theta}F(x)dx. \tag{10}$$

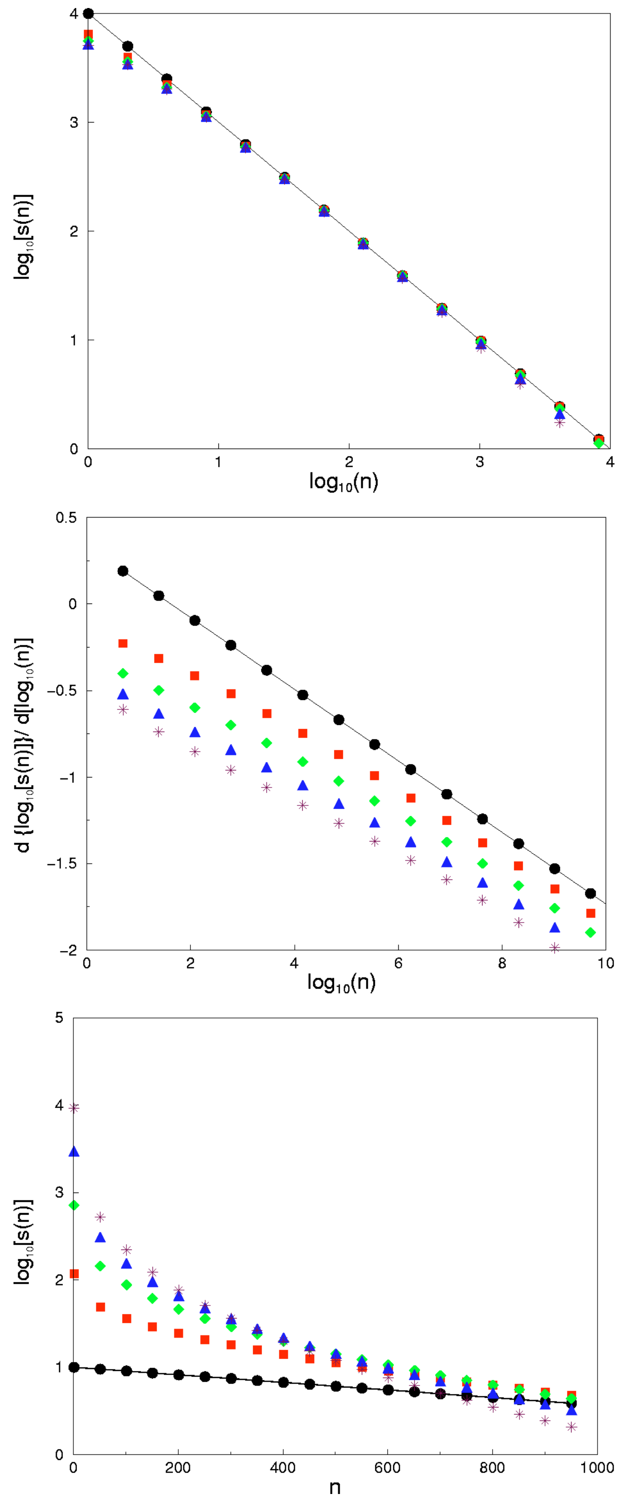

FIG. 2. Plot of the original distribution $s(n)$ and those obtained after the first four iterations, $s^i(n)$, $i=1,\ldots,4$, of the splitting procedure described in Eq. (6). The distributions are depicted by circles, squares, diamonds, triangles, and asterisks, respectively. The three panels refer to the three initial distributions discussed in the text: $s(n)\sim 1/n$ if $1\leqslant n\leqslant n_{max}$, 0 otherwise (power-law, upper panel); $s(n)\sim e^{-[\log_{10}(n/m)]^2/(2t^2)}$ (log normal, middle panel); $s(n)\sim(1-n/n_{max})^{s_{tot}-2}$ (broken stick, lower panel), where m, t, s_{tot}, and n_{max} are constants.

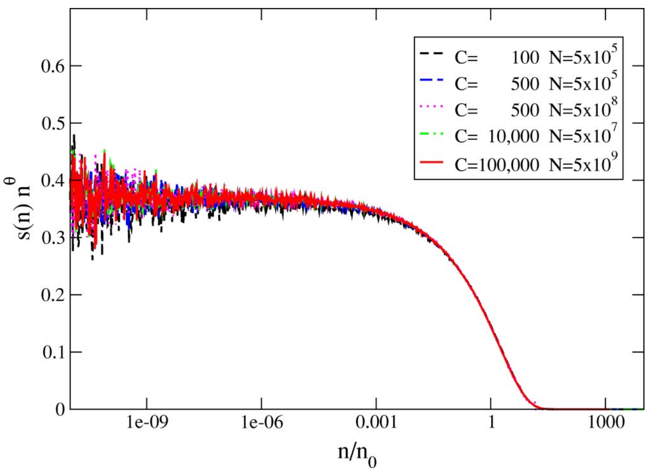

FIG. 3. Scaling collapse plot for the random-split–random-merger case with equal probability for split or merger ($p=1/2$) for differing numbers of companies C and total employees N. For the scaling cutoff n_0 we used the mean company size, N/C. The exponent here is 2/3 as predicted in Eq. (2). We have confirmed that similar scaling, but with different exponents in accord with the theoretical predictions, is found for other values of p and for other split-merger scenarios.

There are two distinct scenarios that one needs to consider. If one requires that the number of companies (species) becomes infinitely large when n_0 and the number of individuals diverge one obtains a more restricted inequality, $\theta \leqslant 1$. This is indeed the case in ecology in which the number of distinct species diverges when the number of individuals diverges. Of course, when there is no such requirement and an infinite number of individuals can be accommodated in a finite number of companies, the weaker inequality for θ holds.

In any case, one finds that n_0 is simply proportional to the average company size (mean species abundance), N/C, and that

$$C \sim N^{(1-\theta)/(2-\theta)}. \tag{11}$$

This leads to the result

$$\Delta C \sim N^{-1/(2-\theta)} \Delta N, \tag{12}$$

which quantifies the increase in the number of companies (species) due to a small increase in the total number of individuals. In a free market society or in an ecosystem in which the biodiversity is maximized, one expects that the impact on the number of companies (species) is minimum when one introduces a small number of additional individuals. This follows from the observation that a large impact on the number of companies (or species) would imply that the economy (ecosystem) is not optimized for maximal profit (biodiversity). On minimizing ΔC subject to the constraints on θ, we find that the optimal value of θ is given by

$$\theta = 1 \quad \text{or} \quad 2. \tag{13}$$

It is remarkable that this drive towards optimality leads to the observed $1/n$ or $1/n^2$ dependencies in the species abundance relationship in ecology and in the company size distribution, respectively. These two distinct universality classes correspond to the exponent being at the edge of the allowed range.

We are indebted to Lord Robert May for reading a preliminary version of this manuscript. This work was supported by NASA, Grant No. Cofin-2001, NSF IGERT Grant No. DGE-9987589, and by the Swiss National Science Foundation under Grant No. FNRS 21-61397.00.

[1] E.W. Montroll and M.F. Schlesinger, Proc. Natl. Acad. Sci. U.S.A. **79**, 3380 (1982).
[2] P. Bak, *How Nature Works: The Science of Self-Organized Criticality* (Springer-Verlag, New York, 1996).
[3] H.E. Stanley, Rev. Mod. Phys. **71**, S358 (1999).
[4] R.N. Mantegna and H.E. Stanley, *Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge University Press, Cambridge, 2000).
[5] A.L. Barabasi, *Linked: The New Science of Networks* (Perseus Press, Cambridge, MA, 2002).
[6] M. Buchanan, *Ubiquity: Why the World Is Simpler Than We Think* (Weidenfeld & Nicholson, London, 2000).
[7] M. Newman, http://www.santafe.edu/mark/powerlaws.pdf, 2002.
[8] D. Sornette, *Critical Phenomena in Natural Sciences* (Springer, Berlin, 2000).
[9] M. Smoluchowski, Phys. Z. **17**, 557 (1916).
[10] J.D. Barrow, J. Phys. A **14**, 729 (1981).
[11] P.L. Krapivsky, I. Grosse, and E. Ben-Naim, Phys. Rev. E **61**, R993 (2000).
[12] P.L. Krapivsky and S.N. Majumdar, Phys. Rev. Lett. **85**, 5492 (2000).
[13] K. Nagel, M. Shubik, M. Paczuski, and P. Bak, Physica A **287**, 546 (2000).
[14] G. Madras and B.J. McCoy, J. Colloid Interface Sci. **246**, 356 (2002).
[15] R.B. Diemer and J.H. Olson, Chem. Eng. Sci. **57**, 2193 (2002).
[16] D. Zheng, G.J. Rodgers, and P.M. Hui, Physica A **310**, 480 (2002).
[17] S. Gueron and S.A. Levin, Math. Biosci. **128**, 243 (1995).
[18] R.L. Axtell, Science **293**, 1818 (2001).
[19] R.M. May, in *Ecology of Species and Communities*, edited by M. Cody and J.M. Diamond (Harvard University Press, Cambridge, MA, 1975), pp. 81–120.
[20] A.K. Dewdney, Math. Intell. **23**, 27 (2001).
[21] R.A. Fisher, A.S. Corbet, and C.B. Williams, J. Anim. Ecol. **12**, 42 (1943).
[22] B. Asfaw *et al.*, Nature (London) **416**, 317 (2002).
[23] A.L. Roca *et al.*, Science **293**, 1473 (2001).
[24] V. Novotny *et al.*, Nature (London) **416**, 841 (2002).
[25] E. Mayr, *Animal Species and Evolution* (Harvard University Press, Cambridge, MA, 1963).