On the Universality of Zipf's Law

Bernat Corominas-Murtra
Ricard V. Solé

SFI WORKING PAPER: 2010-01-002

On the Universality of Zipf's Law

Bernat Corominas Murtra[1] and Ricard Solé[1,2]
[1]ICREA-Complex Systems Lab, Universitat Pompeu Fabra (GRIB), Dr Aiguader 80, 08003 Barcelona, Spain
[2]Santa Fe Institute, 1399 Hyde Park Road, Santa Fe NM 87501, USA

Zipf's law is the most common statistical distribution displaying scaling behavior. Cities, populations or firms are just examples of this seemingly universal law. Although many different models have been proposed, no general theoretical explanation has been shown to exist for its universality. Here we show that Zipf's law is, in fact, an inevitable outcome of a very general class of stochastic systems. Borrowing concepts from Algorithmic Information Theory, our derivation is based on the properties of the symbolic sequence obtained through successive observations over a system with an ubounded number of possible states. Specifically, we assume that the complexity of the description of the system provided by the sequence of observations is the one expected for a system evolving to a stable state between order and disorder. This result is obtained from a small set of mild, physically relevant assumptions. The general nature of our derivation and its model-free basis would explain the ubiquity of such a law in real systems.

Keywords: Kolmogorov Complexity, Entropy, Stochastic System, Zipf's Law

I. INTRODUCTION

Scaling laws are common in both natural and artificial systems (1). Their ubiquity and universality is one of the fundamental issues in statistical physics (2–4). One of the most prominent examples of power law behavior is the so called *Zipf's law* (5–7). It was popularized by the linguist G. K. Zipf, who observed that it accounts for the frequency of words within written texts (5; 8). But this law is extremely common, (9) and has been found in the distribution of populations in city sizes (5; 10–13), firm sizes in industrial countries (14), market fluctuations (15), money income (16), Internet file sizes (17) or family names (18). For instance, if we rank all the cities in a country from the largest (in population size) to the smallest, Zipf's law states that the probability $p(s_i)$ that a given individual lives in the i-th most populated city $(i = 1, ..., n)$ falls off as

$$p(s_i) = \frac{1}{Z} i^{-\gamma}, \qquad (1)$$

with the exponent, $\gamma \approx 1$, and being Z the normalization constant, i.e.,

$$Z = \left(\sum_{i \leq n} i^{-\gamma} \right). \qquad (2)$$

Although systems exhibiting Zipf's-like statistics are clearly different in their constituent units, the nature of their interactions and intrinsic structure, most of them share a few essential commonalities. One is that they are stochastic, far from equilibrium systems changing in time, under mechanisms that prevent them to become homogeneous. Within the context of economic change, for example, wider varieties of goods and attraction for people are fueled by large developed areas. Increasing returns drive further growth and feedback between economy and city sizes (19–21). Moreover, the presence of a scaling law seems fairly robust through time: in spite of widespread political and social changes, the statistical behavior of words in written texts, cities or firms has remained the same over decades or even centuries (5; 7; 14; 21; 22). Such robustness is remarkable, given that it indicates a large insensitivity to multiple sources of external perturbation. In spite of their disparate nature, all seem to rapidly achieve the Zipf's law regime and remain there.

To account for the emergence and robustness of Zipf's law, several mechanisms have been proposed, including auto-catalytic processes (23–25), extinction dynamics (26; 27), intermittency (28; 29), coherent noise (30), coagulation-fragmentation processes (31; 32), self-organized criticality (33), communicative conflicts (34; 35), random *typewriting* (36) or stochastic processes in systems with interacting units with complex internal structure (37). The diverse character of such mechanisms sharing a common scaling exponent strongly points towards the hypothesis that some fundamental property (beyond a given specific dynamical mechanism) is at work. Such a universal trend asks for a generic explanation, which should avoid the use of a particular set of rules.

We address the problem from a very general, mechanism-free viewpoint; by studying the statistical properties of the sequence of successive observations over the system. More precisely, our observations can be understood as a sequence of symbols of a given alphabet (depending on the nature of the system) following some probability distribution. The elements of this alphabet can be coded in some way -for example, bits. From this conceptual starting point, we borrow concepts from algorithmic information theory and propose a characterization of a wide family of stochastic systems, to which those systems displaying Zipf's law would belong, thereby showing that Zipf's law is the only physically relevant solution. Our approach is, to our knowledge, the first general, theoretical explanation to account for the commonality of Zipf's law in so many different systems and contexts.

II. ALGORITHMIC COMPLEXITY OF STOCHASTIC SYSTEMS

A general explanation for the origins of Zipf's law will necessarily require the use of rather fundamental argument. The cornerstone of ours is an abstract characterization of the sequence of observations made on a given system in terms of so called Algorithmic Complexity (38–43) -see also (44). The key quantity of such theory is the so-called *Kolmogorov Complexity*, which is a conceptual precursor of statistical entropy, being a powerful indicator of the complexity (and predictability) of a dynamical system (45–47). In a nutshell, let $\mathbf{x}$ be a symbolic string generated by the successive observations of the system $\mathcal{S}$. Its Kolmogorov Complexity, $K(\mathbf{x})$ is defined as the length $l(\pi^*)$-in bits- of the shortest program π^* executed in a computer in order to reproduce $\mathbf{x}$ (see fig. 1). This measure -which is computer independent, up to an addition constant- has been often used in statistical physics (48–50) particularly in the context of symbolic dynamics (46). In this context, K is known to be maximal for completely disordered -i.e., completely random- systems, whereas it takes intermediate values when some asymmetry on the probabilities of appearance of symbols emerges.

In the following section we outline an abstract characterization of complex systems from the sequence obtained through a set of successive observations. We also briefly discuss some of its implications from the statistical physics viewpoint, using the well-known connection of Kolmogorov complexity with statistical entropy. This connection is exploited to derive Zipf's law as the unique solution of our problem.

A. Stochastic Systems

In the framework of statistical physics, the sequence of observations over a system can be interpreted as a sequence of independent, identically distributed random variables, where the specific outcomes of the observations are obtained according to a given probability distribution -see fig. (1). In mathematical terms, a sequence of observations (obtained from a given system) whose outcome is probabilistic is a *stochastic object*. By definition, the Kolmogorov Complexity of a stochastic object described by a binary string $\mathbf{x} = x_1, .., x_m$ of length m, satisfies the following requirement (51):

$$\lim_{m\to\infty} \frac{K(\mathbf{x})}{m} = \mu \quad \in (0, 1]. \tag{3}$$

In other words, the binary representation of a stochastic object is *linearly* compressible. The case where $\mu = 1$ refers to a completely random object, and the string is called *incompressible*. As an example, let us consider a Bernouilli process, described by a binary random variable X such that $\mathbb{P}(X = 1) = \theta$ $\mathbb{P}(X = 0) = 1 - \theta$, (42). Suppose we perform m observations, thereby generating the sequence of independent, identically distributed random variables

$$X_1, ..., X_m, \tag{4}$$

which is, in this case, a sequence of 1's and 0's. The Kolmogorov Complexity of the string generated by a sequence of m observations over such a stochastic system, $K(X_1, .., X_m)$, satisfies the following scaling relation (51):

$$\lim_{m\to\infty} \frac{K(X_1, ..., X_m)}{m} = \mu; \quad \mu \in (0, 1]. \tag{5}$$

In this case it is straightforward to identify (42)

$$\mu = H(\theta), \tag{6}$$

where $H(\theta)$ is the uncertainty associated to the Bernouilli $\sim \theta$ process, i.e., its *Shannon entropy*:

$$H(\theta) = -\theta \log \theta - (1 - \theta) \log(1 - \theta) \tag{7}$$

(throughout the paper $\log \equiv \log_2$, unless otherwise indicated). The average Kolmogorov Complexity is tied to the uncertainty in predicting, from a given row, the value of the next row, either 1 or 0 -see fig (1a,b). Notice that the most uncertain case is obtained for $\theta = 1/2$, leading to $H(\theta) = 1$, according to the definition of randomness provided above.

We can generalize the concept of random sequence for non binary strings, whose elements belong to a given set $\Sigma = \{s_1, .., s_n\}$, being $|\Sigma| = n$ -see fig (1c,d). This is the case of a dice, for example, whose set of outcomes is $\Sigma_{dice} = \{1, 2, 3, 4, 5, 6\}$. Accordingly, the successive observations of our stochastic system are depicted by a sequence of independent, identically distributed random variables $X_1, ..., X_m$ taking values over the set Σ and following a given probability distribution p. The so-called *noiseless Coding* theorem (42; 52; 53), establishes that the minimum length, (in bits) of the string needed to code the event s_i, $l^*(s_i)$, satisfies

$$l^*(s_i) = -\log(p(s_i)) + \mathcal{O}(1). \tag{8}$$

The average minimum length will correspond to the minimum length of the code, which is, by definition, the Kolmogorov complexity. Thus we obtain the following equality:

$$\lim_{m\to\infty} \frac{K(X_1, ..., X_m)}{m} = \sum_{i\leq n} p(s_i) l^*(s_i) \tag{9}$$

The complete random case is obtained when all the events of Ω are equiprobable, obtaining, for any $s_i \in \Sigma$, $l^*(s_i) = \log n + \mathcal{O}(1)$. This indicates that we need $\approx \log n$

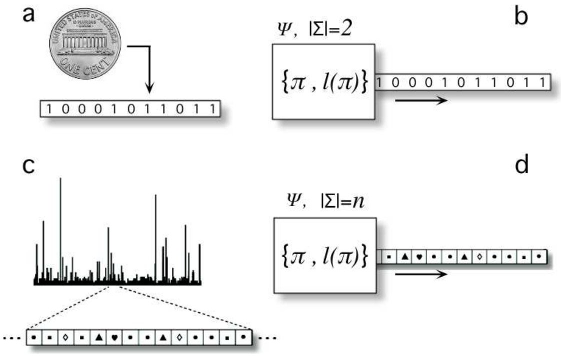

FIG. 1 Algorithmic complexity $K(\mathbf{x})$ of a stochastic string $\mathbf{x}$, indicating a set of observations made on a system, is measured as the lenght in bits of the minimal program p required to reproduce such string. For a fair coin toss (a) which generates a completely random sequence, the computer Ψ (b) would run a program with a length equal to the lenght of the string (which is an upper bound here). Here the size of the alphabet Σ is two (i. e. $|\Sigma| = 2$) but an arbitrary sequence (c) $\mathbf{y}$ obtained from the successive observations over complex system would not be restricted by the binary description. Instead, a large range of n possible symbols would be used to define our string now. This is coded through a minimal program which, when applied to a computer (d), replicates the $n-$ary original sequence. The length of this minimal program, coded in bits, is the Kolmogorov Complexity of $\mathbf{y}$.

bits to code any element from Σ. Therefore, the length in bits of the sequence of m successive observations will be approximately $m \cdot \log n$. If we are not in the special case of equiprobability, it is clear, by observing eq. (8), that the average minimum length of the code will be lower than $\log n$. Using our previous result (5) for the binary case, it is not difficult to see that the analogous of (5) is:

$$\lim_{m\to\infty} \frac{K(X_1, ..., X_m)}{m \cdot \log n} = \mu; \quad \mu \in (0, 1], \tag{10}$$

being μ defined as

$$\mu = -\sum_{i\leq n} p(s_i) \log_n p(s_i), \tag{11}$$

where $\log_n$ is the n-based logarithm. So far we have been concerned with the algorithmic characterization of stochastic systems for which the size of the configuration space is static. However, we must differentiate the properties of the systems we want to characterize from a standard stochastic object such as the ones obtained by tossing a dice or a coin. They both generate a bounded number of possible outcomes -namely, 6 and 2- with an associated probability, whereas those systems exhibiting power-laws lack an a priori constraint on the potential number of available outcomes. These systems are *open* concerning the size -or dimensionality- of the configuration space. To emphasize this property, we introduce an explicit dependence of the random variables on the number of possible outcomes: Let $X(n)$ be a random variable taking values on Σ, where $|\Sigma| = n$. We refer to the probability distribution associated to $X(n)$ as p_n, where (without any loss of generality) an ordering

$$p_n(s_1) \geq p_n(s_2) \geq ... \geq p_n(s_n) \tag{12}$$

is assumed. At a given time, the system satisfies eq. (10), since it is a stochastic object with a given number of available states. However, assuming that the system changes (generally growing) but maintains its basic statistical properties stable (5; 7; 14; 22) . condition (10) is replaced by:

$$\lim_{m,n\to\infty} \frac{K(X_1(n), ..., X_m(n))}{m \cdot \log n} = \mu; \quad \mu \in (0, 1]. \tag{13}$$

We can replace eq. (13) alternatively by the following statement: For any $\epsilon > 0$ there exists $n \in \mathbb{N}$ such that, for any $n' > n$:

$$\left| \lim_{m\to\infty} \frac{K(X_1(n'), ..., X_m(n'))}{m \cdot \log(n')} - \mu \right| < \epsilon. \tag{14}$$

Eqs. (13) and (14) define a scaling relation, indicating that basic statistical properties are invariant under changes in system's size.

A computational test for this result can be illustrated by the model results shown in figure 2. The picture shows a spatial snapshot of the local population densities of a model of urban growth displaying Zipf's law (21). The simulation is performed on a small 80×80 lattice and the normalized entropy evolves towards a stationary value $\mu \approx 0.65$ consistently with our discussion. This is true in spite that this model exhibits wide fluctuations due to its intermittent stochastic dynamics.

Our main objective will be to solve eq. (13). We achieve this objective by exploring two scenarios. First, we propose a power-law probability distribution as a solution, i.e.:

$$p_n(s_i) \propto i^{-\gamma}. \tag{15}$$

This *power-law ansatz* is purely mathematical, and can be replaced by a more physically realistic assumption. This leads us to the second strategy to solve eq. (13), which is based on the assumption that the mechanisms responsible for the growth and stabilization of the system do not depend on the size of the configuration space, and, thus, it is reasonable to assume that a partial observation of the system will satisfy also condition (13). We will refer to this assumption as the *scale invariance condition*, and it is formulated as follows. Let $\Sigma^{(k)} \subseteq \Sigma$ be the set of the first k elements of Σ, observing a labeling consistent with the ordering of probabilities provided in eq. (12) -roughly speaking, the k most probable elements of Σ. The random variable which accounts for the observations of such k elements is notated $X(k \leq n)$. Furthermore, let us define ϵ' as:

$$\epsilon' \equiv \left| \lim_{m\to\infty} \frac{K(X_1(k \leq n), ..., X_m(k \leq n))}{m \cdot \log(k)} - \mu \right| + \delta, \tag{16}$$

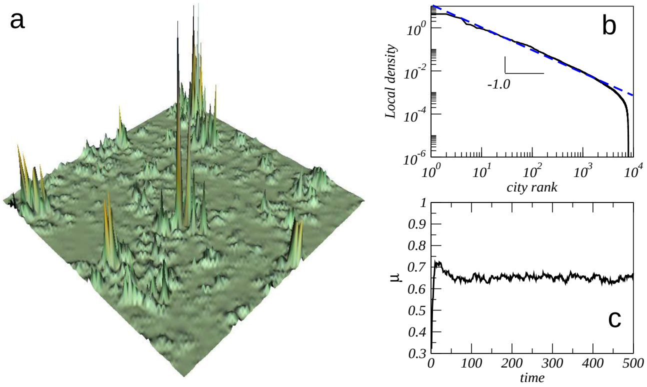

FIG. 2 An example of the behavior of the normalized entropy for a multiplicative stochastic process exhibiting Zipf's law. Here we use the model described in (21) using a 80×80 lattice where each node is described by a density of population $\rho(i,j)$. The rules of the model are very simple: i) At every time step, each node loses a fraction α of its contents, which is distributed among its four nearest neighbors. ii) At time $t+1$ the local population is multiplied, with probability p, by a factor p^{-1}. Furthermore, with probability $1-p$, the population of balls of an urn is set to zero. Here we use $0 < \eta < 0.01$, $\alpha = 1/4$ and $p = 3/4$. This is an extremly simplified (and yet successful) model of urban population dynamics. A snapshot (for $t = 500$) is shown in (a) where we can appreciate the wide range of local densities, following Zipf's law (b). If we plot the evolution of the normalized entropy μ over time (averaged over 10^2 replicas) we observe a convergence towards a stationary value $\mu \approx 0.65$.

being δ arbitrarily small. Then, for any $n \geq k' \geq k$,

$$\left| \lim_{m\to\infty} \frac{K(X_1(k' \leq n), ..., X_m(k' \leq n))}{m \cdot \log(k')} - \mu \right| < \epsilon'. \quad (17)$$

We finally observe that the case where $\mu = 1$ is automatically ruled out, since it belongs to a sequence of outcomes from a completely random system, a category that falls outside the set of systems we are studying. This will become clearer in the next section, where we briefly discuss the implications of the above characterization in terms of statistical physics.

B. Statistical Interpretation

The Kolmogorov Complexity of a sequence of observations over an stochastic system, depicted by a sequence of independent, identically distributed random variables $\mathbf{x} = X_1, ..., X_m$ satisfies the following condition (42; 54):

$$\lim_{m\to\infty} \frac{K(\mathbf{x})}{m} = H(X), \quad (18)$$

where $H(X)$ is the Shannon entropy associated to the random variable X, which takes values on the set $\Sigma = \{s_1, ..., s_n\}$ following the probability distribution p:

$$H(X) = -\sum_{i\leq n} p(s_i) \log p(s_i). \quad (19)$$

This equivalence is the key for the statistical interpretation of eq. (13). Indeed, by defining the *normalized entropy*, $h(n)$ as

$$h(n) \equiv \frac{H(X(n))}{\log n}, \quad (20)$$

eq. (13) is actually analogous to

$$\lim_{n\to\infty} h(n) = \mu; \quad \mu \in (0, 1], \quad (21)$$

and condition (14) can be rewritten as follows: For any $\epsilon > 0$ there exists $n \in \mathbb{N}$ such that, for any $n' > n$,

$$|h(n') - \mu| < \epsilon. \quad (22)$$

To derive the statistical analog of eq. (17), we observe that, if $X(n)$ follows the probability distribution p_n, the k most probable elements of Ω obey the following probability distribution p_n^k:

$$p_n^k(i) \equiv \mathbb{P}(s_i | i \leq k) = \left(\sum_{j\leq k} p_n(s_j) \right)^{-1} p_n(s_i). \quad (23)$$

Thus, if $H(k \leq n)$ is the entropy of the above probability distribution, its normalized counterpart is defined as $h(k \leq n)$:

$$h(k \leq n) \equiv \frac{H(k \leq n)}{\log k}, \quad (24)$$

and condition (17) can be rewritten as follows: Let ϵ' be defined as

$$\epsilon' \equiv |h(k \leq n) - \mu| + \delta, \quad (25)$$

being δ arbitrarily small. Then, for any $n \leq k' \leq k$,

$$|h(k' \leq n) - \mu| < \epsilon'. \quad (26)$$

Consistently, the above conditions on the statistical entropy of our system automatically rule out the trivial, fully random case $\mu = 1$. This particular case imposes equiprobability in the observation of states, and such a symmetry would depict a system in equilibrium where no other constraints than normalization of probabilities are at work, following Jaynes' maximum entropy principle. The case $\mu = 0$ would correspond to systems where, in spite of growing in size, its complexity (and thus, its statistical entropy) is bounded or grows sublinearily with $\log n$, a case studied in (46). Here, we are interested is the intermediate case, where

$$\mu \in (0, 1). \quad (27)$$

This characterization would depict systems with some balance among ordering and disordering forces, and thereby displaying a dissipation of statistical entropy proportional to the maximum entropy achievable for the system in equilibrium. Therefore, we will refer to the problem of finding solutions for eqs. (13, 21) as the *entropy restriction problem.*

Despite the very general character of eq. (13), in this work we demonstrate that, under either condition (15) or (17), if $\mu \in (0,1)$, Zipf's law is the only solution for the probability distribution governing the behavior of $X(n)$.

III. EMERGENCE OF ZIPF'S LAW IN STOCHASTIC SYSTEMS

As pointed out in (34), the main difficulty we face in this kind of equations is that we are not dealing with an extremal problem, since our value of entropy is previously fixed and it is neither minimum nor maximum, in Jaynes' sense (55). Thus, classical variational methods, which have been widely used with great success in statistical mechanics (55–58), do not apply to our problem. We also must take into account that the particular properties of Zipf's law create an additional difficulty if the studied systems display, a priori, an unbounded number of possible states. Specifically, we refer to the non-existence of finite moments and normalization constant in the thermodynamical limit. However, as we shall see, these apparently undesirable properties will be the key to our derivation.

In this section we find the form of the probability distribution compatible with eqs. (13, 21) if $\mu \in (0,1)$, using firstly the power law ansatz depicted in eq. (15) and, secondly, the scale-invariance condition depicted in eqs. (17, 26). Both procedures exploit the special properties of the entropies of a power-law distribution. Let us briefly outline how we will proceed:

1. *Power Law ansatz.*- The first method assumes that the solution of eqs. (13, 21) is a power-law distribution, with an arbitrary exponent $\gamma(n)$, which depends on the size of the system. Then, it is demonstrated that, for any $\alpha > 0$, there is an n, such that, if $n' > n$, then,

$$|\gamma(n') - 1| < \alpha, \tag{28}$$

   which means that, for $n \to \infty$, all exponents will be arbitrarily close to 1, regardless the value of μ.

2. *Scale invariance condition.*- In this second approach we solve eqs. (13, 21) using condition (17, 26), which leads us to the following inequality:

$$\left(\frac{i+1}{i}\right)^{(1-\delta)} > \frac{p_n(i)}{p_n(i+1)} > \left(\frac{i+1}{i}\right)^{(1+\delta)}, \tag{29}$$

   being $\delta \to 0$ as $n \to \infty$.

Thus, from these two approaches we conclude that, at the thermodynamic limit, Zipf's law is the only solution for systems whose sequence of observations satisfy (13), for any $\mu \in (0,1)$.

A. Properties of the entropies of a power law

Let us briefly summarize the properties of the entropies of power-law distributed systems, which will be used to derive the main results of this work -For detailed derivations of this section, see the appendix section. Such properties are intimately linked with the behavior of the *Riemann Zeta function*, $\zeta(\gamma)$ (59):

$$\zeta(\gamma) = \sum_{k=1}^{\infty} \frac{1}{k^{\gamma}}. \tag{30}$$

In the real line, this function is defined in the interval $\gamma \in (1, \infty)$, displaying a singularity at the limit $\gamma \to 1^{+}$.

Now, let us suppose that the system contains n states and the probability to find the i-th most likely states follows a power-law, i.e., $p_n(s_i) \propto i^{-\gamma}$. For the sake of simplicity, we will refer to its associated entropy as $H(n,\gamma)$ and to its normalized counterpart as $h(n,\gamma)$. The most basic properties concern the global behavior of $H(n,\gamma)$. It is straightforward to check that $H(n,\gamma)$ is $i)$ a monotonous increasing function on n_t and $ii)$ a monotonous decreasing function on γ. Concerning more sophisticated features, we first observe that the normalized entropy of Zipf's law of a system with n states (i.e. $p_n(i) \propto i^{-1}$ with $i = 1, ..., n$) converges to 1/2 (60), i.e.,

$$\lim_{n\to\infty} h(n,1) = \frac{1}{2}. \tag{31}$$

We also observe that the entropy of a power law with exponent higher than one is bounded i.e., if $\gamma > 1$ is the exponent of our power law, there exists a finite constant $\phi(\gamma)$ such that:

$$\lim_{n\to\infty} H(n,\gamma) < \phi(\gamma). \tag{32}$$

An interesting consequence of this result is that, if our (unknown) probability distribution is dominated[1] from some k by some power-law with exponent $\gamma > 1+\delta$ (for any $\delta > 0$), our entropy will be bounded.

Furthermore, it can be shown that the normalized entropy of a power-law distribution in a system with n different states, with exponent $\gamma < 1$, converges to 1, i.e.,

$$\lim_{n\to\infty} h(n,\gamma) = 1. \tag{33}$$

[1] A probability distribution is dominated from some k by a power law with exponent $1+\delta$ if $(\exists m) : (\forall i > m)\left(\frac{p(i+1)}{p(i)} < \left(\frac{i}{i+1}\right)^{1+\delta}\right)$.

Consistently, we can conclude that, if an (unknown) probability distribution is not dominated from any m by a power law with exponent lower than $1-\delta$ (for any $\delta > 0$), the normalized entropy of our system will converge to 1.

Using these properties, in the following sections we proceed to derive Zipf's law starting from the previous assumptions.

B. Power Law Ansatz: Convergence of Exponents to $\gamma = 1$

In this section we make use of the power-law ansatz as a solution of our problem -eq. (15). Let us rewrite the convergence assumptions provided in (14, 21) assuming that our probability distribution is a power law with unknown exponent γ: For any $\epsilon > 0$ we can find an n such that, for any $n' > n$,

$$|h(n', \gamma(n')) - \mu| < \epsilon, \tag{34}$$

i.e., the sequence of normalized entropies $\mathcal{H}$, associated to system's growth, namely

$$\mathcal{H} = h(1, \gamma(1)), h(2, \gamma(2)), ..., h(k, \gamma(k)), ..., \tag{35}$$

converges to μ. Below we split the problem in two different scenarios.

1. First case: $\mu < \frac{1}{2}$.

We begin by exploring the following scenario:

$$\lim_{n\to\infty} h(n, \gamma(n)) = \mu \in \left(0, \frac{1}{2}\right). \tag{36}$$

From equation (31) we can ensure that, for large values of n, $\gamma(n) > 1$. Since we assumed that the sequence $\mathcal{H}$ converges to μ, we can state that, for a given $\epsilon > 0$, there is an arbitrary n_1 such that:

$$\mu - \epsilon < h(n_1, \gamma(n_1)) < \mu + \epsilon. \tag{37}$$

We know, from the properties of the entropies of power-law distributed systems, that $H(n_1, \gamma(n_1)) < \phi(\gamma(n_1))$, where $\phi(\gamma(n_1))$ is some positive, finite constant (see eq. (32) and appendix). Then, since $\log x$ is an unbounded, increasing function of x, we can find $n_2 > n_1$ such that

$$\phi(\gamma(n_1)) < (\mu + \epsilon) \log n_2. \tag{38}$$

Thus, since $h(n, \gamma)$ is a decreasing function on γ, we need to find $\gamma(n_2) < \gamma(n_1)$ such that

$$\mu - \epsilon' < h(n_2, \gamma(n_2)) < \mu + \epsilon', \tag{39}$$

with $\epsilon' \leq \epsilon$, in order to satisfy the entropy restriction. Furthermore, since $H(n, 1) = \frac{1}{2} \log n + \mathcal{O}(\log(\log n))$, we conclude that $1 < \gamma(n_2) < \gamma(n_1)$. Let us expand this process recursively, thus generating an infinite decreasing sequence of exponents,

$$\{\gamma(n_k)\}_{k=1}^{\infty} = \gamma(n_1), ..., \gamma(n_i), ..., \tag{40}$$

such that, for any $\gamma(n_i) \in \{\gamma(n_k)\}_{k=1}^{\infty}$, $\gamma(n_i) > 1$. We observe that, for any $\alpha > 0$, we can find a n_k such that, if $n_j > n_k$,

$$|\gamma(n_j) - 1| < \alpha, \tag{41}$$

since, for every $\gamma(n_k)$, we always find a $n_j > n_k$ such that

$$\phi(\gamma(n_k)) < (\mu + \epsilon) \log n_j. \tag{42}$$

2. Second case: $\mu > \frac{1}{2}$.

Let us now consider the following entropy restriction problem:

$$\lim_{n\to\infty} h(n, \gamma(n)) = \mu \in \left(\frac{1}{2}, 1\right). \tag{43}$$

From equation (31), we can ensure that, for any n, $\gamma(n) < 1$. Furthermore, from equation (33), we again find a problem close to the one solved above, since for n_1 large enough and $\gamma < 1$, we have:

$$H(n_1 + 1, \gamma) - H(n_1, \gamma) > \mu(\log(n_1 + 1) - \log n_1). \tag{44}$$

Now, since we assumed that the sequence $\mathcal{H}$ converges, we can state that given an arbitrary step n_1,

$$\mu - \epsilon < h(n_1, \gamma(n_1)) < \mu + \epsilon. \tag{45}$$

Since $H(n, \gamma)$ is a decreasing function on γ, we need to find $\gamma(n_2) > \gamma(n_1)$ such that:

$$\mu - \epsilon' < h(n_2, \gamma(n_2)) < \mu + \epsilon', \tag{46}$$

with $\epsilon' \leq \epsilon$, to satisfy the entropy restriction. However, from equation (31), we know that $1 > \gamma(n_2) > \gamma(n_1)$. Proceeding as above, we expand this process, thus generating an infinite increasing sequence of exponents $\{\gamma(n_k\}_{k=1}^{\infty}$. By virtue of equation (31) and equation (33), and taking into account the decreasing behavior of h as a function of the exponent, we observe that, for any $\alpha > 0$, we can find a n_k such that, if $n_j > n_k$,

$$|\gamma(n_j) - 1| < \alpha. \tag{47}$$

In summary, under the power law ansatz, the only solution for eqs. (13, 21), in the limit of large systems, is $\gamma = 1$, i.e., Zipf's law. An illustrative picture of the above reasoning is provided in fig. (3) in the appendix, where it becomes explicit that, when computing the normalized entropy of a power law, the larger the size of the system, the bigger is the interval of normalized entropies whose exponent is close to 1.

C. Scale invariance Condition

This condition, depicted in eqs. (17, 26), is grounded on the assumption that the entropy restriction works at all levels of observation. Thus, the partial probability distributions of states we obtain must reflect the effect of the entropy restriction, introducing a *scale invariance* of the normalized entropy of the partial samples of the system.

For the sake of clarity, let us rewrite the main equations of this scaling argument. The key point is that the scale invariance of the entropy restriction assumption implies that the normalized entropy of our system converges in the following way -see eqs. (23, 24, 17, 26): Let ϵ' be defined as

$$\epsilon' \equiv |h(k \leq n) - \mu| + \delta, \tag{48}$$

being δ arbitrarily small. Then, for any $n \leq k' \leq k$:

$$|h(k' \leq n) - \mu| < \epsilon'. \tag{49}$$

The consequence of this condition is that the tail of the distribution p_n must be able to unboundedly increase the entropy of the whole system to reach the global value $H(X(n))$, which lies in the interval $((\mu - \epsilon)\log n, (\mu + \epsilon)\log n)$. However, as we saw in the above sections, the decay of this tail is strongly constrained by the entropy restriction, since only special cases avoid the normalized entropy to fall to 0 or 1. On one hand, for any δ and for large i's,

$$\frac{p_n(s_i)}{p_n(s_{i+1})} < \left(\frac{i+1}{i}\right)^{(1-\delta)} \tag{50}$$

to avoid that $h(n) \to 1$. On the other hand, if we want to avoid that $h(n) \to 0$, the following inequality must hold:

$$\frac{p_n(s_i)}{p_n(s_{i+1})} > \left(\frac{i+1}{i}\right)^{(1+\delta)}. \tag{51}$$

Thus, the solution of our problem lies in the range defined by:

$$\left(\frac{i+1}{i}\right)^{(1-\delta)} > \frac{p_n(s_i)}{p_n(s_{i+1})} > \left(\frac{i+1}{i}\right)^{(1+\delta)}. \tag{52}$$

From the study of the entropies of a power law performed in the previous section, we know that δ can be arbitrarily small if the size of the system is large enough; thus,

$$\frac{p_n(s_i)}{p_n(s_{i+1})} \approx \frac{i+1}{i} \;\Rightarrow\; p_n(s_i) \propto i^{-1}, \tag{53}$$

which leads us to Zipf's law as the unique asymptotic solution.

IV. DISCUSSION

Complex, far from equilibrium systems involve a tension between amplifying mechanisms and negative feedbacks able to buffer the impact of fluctuations. In this paper we have considered the consequences of such tension in terms of one of its most well known outcomes: the presence of an inverse scaling law connecting the size of observed events and its rank. The commonality of Zipf's law in both natural and man-made systems has been a puzzle that attracted for years the attention of scientists, sociologists and economists alike. The fact that such a plethora of apparently unrelated systems display the same statistical pattern points towards some fundamental, unifying principle.

In this paper we treat complex systems as stochastic systems describable in terms of algorithmic complexity and thus statistical entropy. A general result from the algorithmic complexity theory is that eq. (3) holds for stochastic systems. Taking this general result as the starting point, we define a characterization of a wide class of complex systems which grasps the open nature of many complex systems, summarized in eq. (21). The main achievement of this equation is that it encodes the concepts of growing and, even most important, the stabilization of complexity properties in an intermediate point between order and disorder, a feature observed in many systems displaying Zipf's-like statistics. From the statistical analog of this equation we derived Zipf's law as the natural outcome of systems belonging to this class of stochastic systems.

Our development avoids the classical procedures based on maximization (minimization) of some functional in order to find the most probable configuration of states, since far from equilibrium the ensemble formalism, together with Jaynes' maximum entropy principle (55) can fail due to the open, non-reversible behavior of the systems considered here. Thus we do not introduce moment constraints, as it is usual in equilibrium statistical mechanics (58), but instead a constraint on the value achieved by the normalized entropy, no matter the scale we observe the system. Both a scaling ansatz and a more general scale invariance assumption lead to Zipf's law as the unique solution for this problem.

Acknowledgments

We thank D. Jou, S. Manrubia, J. Fortuny and our colleagues at the Complex Systems Lab for their useful comments. We also acknowledge the helpful comments provided by two anonymous referees. This work has been founded by the McDonnell Foundation (BCM) and by the Santa Fe Institute (RS).

References

[1] Stanley, H-E, Amaral, L-A-N, Gopikrishnan, P, Ivanov, P-Ch, Keitt, T-H, Plerou, V. (2000) *Physica A* **281**, 60-68
[2] Stanley, H-E (1999) *Rev Mod Phys.* **71**, S358-S366.
[3] Newman, M-E-J (2005) *Contemporary Physics* **46** 323-351.
[4] Solé, R-V, Goodwin, B (2001) *Signs of Life: How Complexity Pervades Biology.* (New York, NY. Basic Books).
[5] Zipf, G-K (1949) *Human Behavior and the Principle of Least Effort.* (Addison-Wesley (Reading MA)).
[6] Auerbach, F, (1913) *Patermans Geograpische Mittelungen* **59**, 74-76.
[7] Gabaix, X (1999) *The Quart J of Economics* **114**, 739-767.
[8] Ferrer-i-Cancho, R, Sole, R-V (2002) *Adv in Complex Sys* **5**, 1-6.
[9] For a very complete collection of the works concerning the topic we refer the reader to: http://www.nslij-genetics.org/wli/zipf/.
[10] Simon, H-A (1955) *Biometrika* **42**, 425-440.
[11] Makse, H-A, Havlin, S, Stanley, H-E (1995) *Nature* **377**, 608-612.
[12] Krugman, P (1996) *J Jap Int Econ* **10**, 399-418.
[13] Blank, A, Solomon, S (2000) *Phys A: Stat Mech and its Appl* **287**, 279-288.
[14] Axtell, R L (2001) *Science* **293**, 1818-1820.
[15] Gabaix, X, Gopikrishnan, P, Plerou, V, Stanley, E H (2003) *Nature* **423**, 267-270.
[16] Pareto, V. (1896) *Cours d'Economie Politique,* (Droz. Geneva).
[17] Reed, W J, Hughes, B D (2002) *Phys Rev E* **66**, 067103-067106.
[18] Zanette, D-H, Manrubia, S-C (2001) *Physica A: Stat Mech and its Appl* **295**, 1-8.
[19] Krugman, P (1996) *The Self-Organizing Economy.* (Blackwell. Oxford).
[20] Arthur, B-W (1996) *Increasing returns and path dependence in economy.* (Michigan University Press. Michigan).
[21] Manrubia, S-C, Zanette, D-H, Solé, R-V (1999) *Fractals* **7**, 1-8.
[22] Rozman, G (1990) *East asian Urbanization in the nineteenth century: Comparisons with Europe* ed der Woude et al, V (Clarendon Press), pp. 61-63.
[23] Solomon, S, Levy, M (1996) *Int J Mod Phys C* **7**, 745-751.
[24] Malcai, O, Biham, O, Solomon, S (1999) *Phys Rev E* **60**, 1299-1303.
[25] Huang, Z-F Solomon S (2000) cond-mat/0008026.
[26] Solé, R-V, Manrubia, S-C (1996) Phys Rev E **54**, R42-R45
[27] Newman M-E-J, Palmer, R-G (2003) *Modelling Extinction* (Oxford University Press: New York)
[28] Zanette, D-H, Manrubia, S-C (1997) *Phys Rev Lett* **79**, 523-526.
[29] Manrubia, S-C, Zanette, D-H (1998) *Phys Rev E* **58**, 295-302.
[30] Newman, M-E-J, Sneppen, K (1996) *Phys Rev E* **54**, 6226-6231.
[31] White, W-H (1982) *J Colloid Interface Sci* **87**, 204-208.
[32] Family, F, Meakin, P (1989) *Phys Rev A* **40**, 3836-3854.
[33] Bak, P, Tang, C, Wiesenfeld, K (1987) *Phys Rev Lett* **59**, 381–384.
[34] Harremoës, P, Topsoe, F (2001) *Entropy* **3**, 191-226.
[35] Ferrer-i-Cancho, R, Solé, R V (2003) *Proc Natl Acad Sci USA* **100**, 788-791.
[36] Li, W (1992) *IEEE Trans Inform Theo* **38**, 1842-1845.
[37] Amaral, L-A-N, Buldyrev, S-V, Havlin, S, Salinger, M-A, Stanley, H-E (1997) *Phys Rev Lett* **80** 1385-1388.
[38] Solomonoff, R. (1964) *Inform and Control* **7-1**, 1-22.
[39] Kolmogorov, A (1965) *Problems Inform Transmission* **1**, 1-7.
[40] Chaitin, G-J (1966) *J ACM* **13**, 547-569.
[41] Ming, L, Vitányi, P (1997) *An introduction to Kolmogorov complexity and its applications.* (Springer, New York [u.a.]).
[42] Cover, T-M, Thomas, J-A (1991) *Elements of Information Theory.* (John Wiley and Sons. New York).
[43] Adami, C. (1999) *Introduction to Artificial Life* (Springer, New York)
[44] Chaitin, G-J (1975) *J Assoc Comput Mach* **22**, 329-340.
[45] Nicolis, J-S (1986) *Rep Prog Phys* **49**, 1109-1196.
[46] Ebeling, W, Nicolis, G (1991) *Europhys Lett* **14(3)**, 191-196.
[47] Ebeling, W (1993) *Physica A: Statistical Mechanics and its Applications* **194**, 563-575.
[48] Kaspar, F, Schuster, H-G (1987) *Phys Rev A* **36(2)**, 842-848.
[49] Zurek, W-H (1989) *Nature* **341**, 119-124.
[50] Dewey, T-G (1996) *Phys Rev E* **54**, R39-R41.
[51] Grünwald, P-D, Vitányi, P-M (2008) *Handbook of the Philosophy of Science, Volume 8: Philosophy of Information.*, eds. Adriaans, P, van Benthem, J (Elsevier Science Publishers), pp. 289–325.
[52] Shannon, C-E (1948) *Bell System Technical Journal* **27**, 379-423.
[53] Ash, R-B (1990) *Information Theory.* (New York. Dover).
[54] Grünwald, P-D, Vitányi, P-M (2003) *J Logic, Lang Inform* **12**, 497-529.
[55] Jaynes, E-T (1957) *Phys Rev* **106**, 620-630.
[56] Haken, H (1978) *Synergetics: An Introduction. Nonequilibrium Phase Transitions and Self- Organization in Physics, Chemistry and Biology (Springer Series in Synergetics).* (Springer).
[57] Kapur, J-N (1989) *Maximum-entropy Models in Science and Engineering.* (New Delhi. Wiley Eastern Limited).
[58] Pathria, R-K (1996) *Statistical Mechanics, Second Edition.* (Butterworth-Heinemann).
[59] Abramowitz, M, Stegun, I (editors) (1965) *Handbook of mathematical functions*, NBS, Appl. Math. Ser. (U.S. Government Printing office, Washington, D.C.) Vol. 55.
[60] Jones, D-S (1979) *Elementary Information Theory.* (Oxford University Press. Oxford).

APPENDIX A: Entropic Properties of Power-Law distributed systems

Let us suppose we have a system whose behavior is described by the random variable $X(n)$ taking values on the set

$$\Omega = \{s_1, ..., s_n\}; \quad |\Omega| = n \tag{A1}$$

according to the probabilty distribution $p_n(s_i)$. The labeling 'i' of the state is chosen in such a way that:

$$p_n(s_1) \geq p_n(s_2) \geq ... \geq p_n(s_i) \geq ... \geq p_n(s_n). \quad \text{(A2)}$$

The Shannon entropy of our system of n states, to be noted $H(X(n))$, is defined as (53):

$$H(X(n)) = -\sum_{k\leq n} p_n(s_k) \log p_n(s_k). \quad \text{(A3)}$$

Knowing that:

$$\max\{H(X(n))\} = \log n, \quad \text{(A4)}$$

we define the normalized entropy of the system, to be noted, for simplicity, $h(n)$, as:

$$h(n) \equiv \frac{H(X(n))}{\log n}. \quad \text{(A5)}$$

We will work with power-law distributions:

$$p_n(s_i) = \frac{1}{Z} i^{-\gamma}, \quad \text{(A6)}$$

where n is the number of available states, and Z the normalization constant, which depends on the size of the system, n. Let us rewrite the function $H(X(n))$ as a function of the exponent and the size of Ω, $H(n,\gamma)$. Consistently,

$$h(n,\gamma) \equiv \frac{H(n,\gamma)}{\log n}. \quad \text{(A7)}$$

This appendix is devoted to derive five properties of the entropy of power-law distributed systems:

1. The entropy of a power law is a continuous, decreasing function on the exponent.

2. The entropy of a power-law is a monotonous, increasing function on the size of the system.

3. The normalized entropy of Zipf's law of a system with n states ($p_n(s_i) \propto i^{-1}$) converges to 1/2:
$$\lim_{n\to\infty} \frac{H(n,1)}{\log n} = \frac{1}{2}. \quad \text{(A8)}$$

4. The Entropy of a power law with exponent higher than 1 is bounded.
Consequence: If an unknown probability distribution is dominated from some k by some power-law with exponent higher than $1+\delta$, our entropy will be bounded.

5. The normalized entropy of a power-law distribution in a system with n different states, p_n with exponent lower than 1 converges to 1:
$$\lim_{n\to\infty} \frac{h(n,\gamma)}{\log n} = 1. \quad \text{(A9)}$$
Consequence: If our (unknown) probability distribution is not dominated from some k by a power law with exponent higher than $1-\delta$, our normalized entropy will converge to 1.

1. The qualitative behavior of the Entropy as a function of the size and the exponent

In this section we study and derive *properties* 1 and 2.

1. $H(n,\gamma)$ is a continuous, monotonous decreasing function with respect to γ in the range $(0,\infty)$. Indeed, the dominant term of its derivative is:

$$\frac{\partial H(n,\gamma)}{\partial \gamma} \sim -\sum_{i\leq n} \frac{(\log i)^2}{i^\gamma} < 0. \quad \text{(A10)}$$

2.This property is not so straightforward[2]. We want to show that $H(n,\gamma)$ is a monotonous increasing function on n. In order to prove it, we must compute the difference $H(n+1,\gamma) - H(n,\gamma)$. For simplicity, let us define:

$$S_n \equiv \sum_{k\leq n} \frac{1}{k^\gamma}. \quad \text{(A11)}$$

Using the trivial inequality:

$$\log\left(S_n + \frac{1}{(1+n)^\gamma}\right) > \log(S_n), \quad \text{(A12)}$$

we can state that:

$$H(n+1,\gamma) - H(n,\gamma) = \frac{\gamma}{S_n + \frac{1}{(1+n)^\gamma}} \sum_{k\leq n+1} \frac{\log k}{k^\gamma} + \log\left(S_n + \frac{1}{(1+n)^\gamma}\right) - \frac{\gamma}{S_n} \sum_{k\leq n} \frac{\log k}{k^\gamma} + \log(S_n)$$

[2] The reader could object that this section is unnecessary, since the axiomatic derivation of the uncertainty function (which we call entropy) assumes that the entropy increases with the size of the system. However, the explicit statement of this axiom corresponds to the special case of uniform probabilities (53). Specifically, the axiom states that, *if we have two systems A, B such that A contains n states $a_1, ..., a_n$ and B contains $n+1$ states, $b_1, ..., b_{n+1}$, then, if $(\forall i \leq n) p(a_i) = 1/n$ and* $(\forall i \leq n+1) p(b_i) = 1/(n+1)$ $H(A) < H(B)$.Thus, if we are not dealing with this special case, we need to explicitly demonstrate that it holds for our purposes.

$$
\begin{aligned}
&> \gamma \sum_{k\le n} \frac{\log k}{k^\gamma}\left(\frac{1}{S_n + \frac{1}{(n+1)^\gamma}} - \frac{1}{S_n}\right) + \gamma \frac{\log(n+1)}{S_n + \frac{1}{(n+1)^\gamma}}\\
&= \frac{\gamma}{S_n^2 (n+1)^\gamma + S_n}\left(S_n (n+1)^\gamma \log(n+1) - \sum_{k\le n} \frac{\log k}{k^\gamma}\right)\\
&> 0.
\end{aligned}
$$

Finally, it is easy to check that the following properties also hold:

$$
\begin{aligned}
\lim_{\gamma\to\infty} H(n,\gamma) &= 0, &\text{(A13)}\\
\lim_{\gamma\to 0} H(n,\gamma) &= \log n. &\text{(A14)}
\end{aligned}
$$

2. Asymptotic values of the normalized entropy

This section is devoted to study and derive *properties* $3, 4, 5$. Following the above section,

$$
h(n,\gamma) \equiv \frac{H(n,\gamma)}{\log n}. \tag{A15}
$$

3. We assume that our distribution follows a power-law:

$$
p_n(s_i) = \frac{1}{Z} i^{-1}, \tag{A16}
$$

where Z is the normalization constant. We want to show that the sequence

$$
\mathcal{H} = \{h(k,1)\}_{k=1}^{\infty} = h(1,1), h(2,1), ..., h(k,1), ... \tag{A17}
$$

converges to $\frac{1}{2}$. Let us suppose that $\mathcal{H}$ is a sequence satisfying the above requirements. Then, the entropy for a given n can be approached by (60):

$$
H(n,1) = \frac{1}{2}\log n + \mathcal{O}(\log(\log n)). \tag{A18}
$$

Thus, if $h(n,1) = H(n,1)/\log n$, let us define $\epsilon(n)$ like:

$$
\epsilon(n) \equiv \left| h(n,1) - \frac{1}{2}\right| = \left| \frac{\mathcal{O}(\log(\log n))}{\log n}\right|. \tag{A19}
$$

Clearly, $\epsilon(n)$ is strictly decreasing on n, and, furthermore,

$$
\lim_{n\to\infty} \epsilon(n) = 0. \tag{A20}
$$

4. Here we demonstrate that the entropy of a power law with exponent higher than 1 is bounded[3]. Specifically, we assume there exists a pair of positive constants Z, δ, such that:

$$
p_n(i) = \frac{1}{Z} i^{-(1+\delta)} \tag{A21}
$$

Then, the sequence of $\mathcal{H} = \{h(k, 1+\delta)\}_{k=1}^{\infty}$ converges to 0. Indeed, let us first note that:

$$
\lim_{n\to\infty} p_n(s_i) = \frac{1}{\zeta(1+\delta)} i^{-(1+\delta)}, \tag{A22}
$$

where

$$
\zeta(1+\delta) \equiv \sum_{k}^{\infty} \frac{1}{k^{1+\delta}} \tag{A23}
$$

is the Riemann zeta-function (59). The function is defined by an infinite sum which converges, in the real line, if $\delta > 0$, i.e.:

$$
\sum_{k}^{\infty} \frac{1}{k^{1+\delta}} < \infty. \tag{A24}
$$

otherwise, the sum diverges. Furthermore, it is also true that the above condition also holds for the following series:

$$
\sum_{k}^{\infty} \frac{\log k}{k^{1+\delta}}. \tag{A25}
$$

Indeed, note that, given an arbitrary $\delta > 0$ there exists a finite number i^* such that:

$$
i^* \equiv \min\left\{ i : \left(\delta - \frac{\log(\log i)}{\log i}\right) > 0\right\} \tag{A26}
$$

and, if we define the following exponent, $\beta(i^*)$:

$$
\beta(i^*) \equiv 1 + \delta - \frac{\log(\log i^*)}{\log i^*}, \tag{A27}
$$

there exists a finite constant, $\Psi(\delta)$, defined as:

$$
\Psi(\delta) \equiv \sum_{i<i^*}\left(\frac{\log i}{i^{1+\delta}} - \frac{1}{i^{\beta(i^*)}}\right) + \zeta(\beta(i^*)), \tag{A28}
$$

such that:

$$
\sum_{k}^{\infty} \frac{\log k}{k^{1+\delta}} < \Psi(\delta). \tag{A29}
$$

[3] This derivation is equivalent to the one found in (34), Theorem 8.2. In this theorem, the authors demonstrate that every infinite distribution with infinite entropy is *hyperbolic*, which implies that the distribution is not dominated by a power law with an exponent higher than 1.

With the above properties, it is clear that, if there exists a constant $\phi(1+\delta) < \infty$ such that:

$$\lim_{n\to\infty} H(n, 1+\delta) < \phi(1+\delta), \tag{A30}$$

then, the entropy of a power law with exponent higher than 1 is bounded. As we shall see, it is straightforward by checking directly the behavior of $H(n, 1+\delta)$:

$$\lim_{n\to\infty} H(n, 1+\delta) = \frac{1+\delta}{\zeta(1+\delta)} \sum_{i=1}^{\infty} \frac{\log i}{i^{1+\delta}} + \log(\zeta(1+\delta)).$$

Since $H(n, \gamma)$ is an increasing function on n, and

$$\frac{1+\delta}{\zeta(1+\delta)} \sum_{i=1}^{\infty} \frac{\log i}{i^{1+\delta}} + \log(\zeta(1+\delta)) < \infty, \tag{A31}$$

we can define a constant $\phi(1+\delta)$,

$$\phi(1+\delta) \equiv \lim_{n\to\infty} H(n, 1+\delta) + \epsilon \tag{A32}$$

(where ϵ is any positive, finite constant). Clearly,

$$H(n, 1+\delta) < \phi(1+\delta). \tag{A33}$$

Thus,

$$\begin{aligned} \lim_{n\to\infty} h(n, 1+\delta) &= \lim_{n\to\infty} \frac{H(n, 1+\delta)}{\log n} \\ &\leq \lim_{n\to\infty} \frac{\phi(1+\delta)}{\log n} \\ &= 0. \end{aligned}$$

The above property has an interesting consequence. Let us assume that there exists a value $k < n$ such that, if $i > k$, $\forall \delta > 0$:

$$\frac{p_n(s_i)}{p_n(s_{i+1})} > \left(\frac{i+1}{i}\right)^{(1+\delta)}. \tag{A34}$$

In other words: from some point of the distribution, it decreases faster than a power law with exponent higher than 1. If this is the case,

$$\lim_{n\to\infty} h(n) = 0. \tag{A35}$$

5. This property refers to the case when the exponent is lower than 1. As we shall see, in this case, $h(n, \gamma) \to 1$. Indeed, let us suppose that we have the following probability distribution, with $0 < \delta < 1$:

$$p_n(s_i) = \frac{1}{Z} i^{-(1-\delta)}. \tag{A36}$$

Note that (60):

$$\sum_{k\leq n} \frac{1}{k^{1-\delta}} = \int_1^n \frac{1}{x^{1-\delta}} + \mathcal{O}(1) \approx \frac{n^\delta}{\delta}. \tag{A37}$$

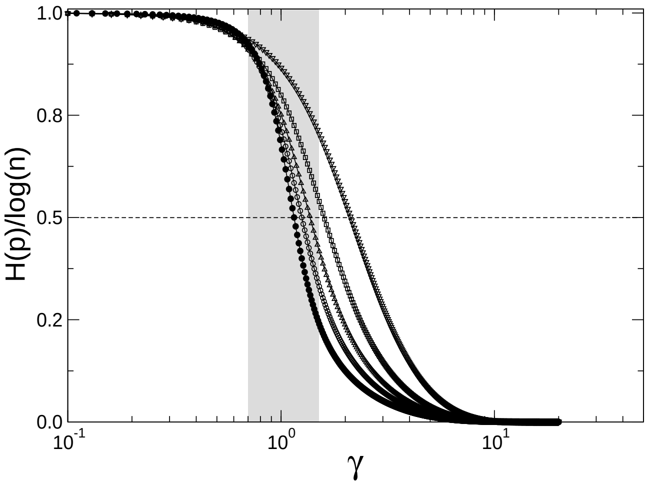

FIG. 3 Normalized entropies of five power-law distributed systems of different size as functions of the exponent. The curves display 5 different sizes. For black circles: $n = 500000$, for white circles: $n = 10000$, for up triangles: $n = 10000$, for squares: $n = 1000$ and, finally, for down triangles: $n = 100$.

Applying directly the definition of entropy,

$$H(n, 1-\delta) = \frac{\delta(1-\delta)}{n^\delta} \sum_{k\leq n} \frac{\log k}{k^{1-\delta}} + \delta \log n - \log \delta. \tag{A38}$$

If we compute the limit of $h(n, 1-\delta) = H(n, 1-\delta)/\log n$:

$$\begin{aligned} \lim_{n\to\infty} h(n, 1-\delta) &= \lim_{n\to\infty} \left(\frac{\delta(1-\delta)}{\log n \cdot n^\delta} \sum_{k\leq n} \frac{\log k}{k^{1-\delta}} + \delta \right) \\ &= \lim_{n\to\infty} \frac{1-\delta}{\log n} \left(\log n - \frac{1}{\delta} \right) + \delta \\ &= 1 - \delta + \delta \\ &= 1. \end{aligned}$$

Again, we can extract an interesting consequence, absolutely symmetrical to the one derived from **4**. Indeed, let us suppose that there exists $k < n$ such that, if $i > k$, $\forall \delta > 0$:

$$\frac{p_n(s_i)}{p_n(s_{i+1})} < \left(\frac{i+1}{i}\right)^{(1-\delta)}. \tag{A39}$$

In other words that, from some point of the distribution, it decreases slower than any power law with exponent lower than 1. If this is the case,

$$\lim_{n\to\infty} h(n) = 1. \tag{A40}$$

3. Numerical values of h as a function of size and exponent

In the section named *Convergence of exponents to* 1 of the main text, we provide a mathematical argument to demonstrate that, if we propose a power-law to solve the following equation (equation (13) of the main text):

$$\lim_{n\to\infty} h(n,\gamma) = \mu \in (0,1), \tag{A41}$$

the exponent of the solution lies arbitrarily close to 1, no matter the explicit value of μ, if it belongs to the open interval $(0,1)$.

In this section we numerically analyze the normalized entropy of a power-law as a function of its size n (number of available states) and exponent, γ, $h(n,\gamma)$:

$$h(n,\gamma) = \frac{1}{\log n}\left(\gamma Z \sum_{i=1}^{n} \frac{\log i}{i^{\gamma}} + \log\frac{1}{Z}\right). \tag{A42}$$

Consistently with the mathematical results, the most interesting feature of the numerical computations is the sharp decay of the normalized entropy when the values of the exponent are about 1, which implies that a wide range of normalized entropies are obtained tuning the exponent of the power-law distribution close to 1. Furthermore, we observe that the decay is sharper if the size of the system grows, concentrating an increasing range of relative entropies near the exponent 1 (grey area). Such a property is used in the main text to demonstrate that, if we make the power law ansatz to solve the problem formulated in equation (13) of the main text, the only solution in the limit of very large systems is the Zipf's law.