geographical analysis

Geographical Analysis (2022) **0**, n/a

City Size Distribution Analyses Based on the Concept of Entropy Competition

Antonio Sanchirico, Giovanna Andrulli, and
Mauro Fiorentino

Department of European and Mediterranean Cultures: Architecture, Environment and Cultural Heritage (Dicem), University of Basilicata, Matera, Italy

The present work pursues theoretical and empirical objectives. With regards to the former, it is demonstrated that the natural tendency to uniformity of both the probability distribution of a city to have a certain number of inhabitants and that of a person to reside in a town of a given number of citizens leads to a competition between their information entropies, which provides the power law distribution as the most probable one for city size. It is also shown that Zipf's law reflects the significant control of the existence of interconnections between cities on the self-organization of their size. With regards to the empirical objectives, based on population data of European countries and Italian municipalities, the theoretical approach proposed is validated. At the Italian scale, city distribution is shown to be a power law for cities above 10,000 inhabitants. In the 20 Italian regions, the breakpoint in the distribution is generally lower. Finally, the geographical control on city distribution is discussed based on the results achieved in some regions.

Introduction

Since Zipf (1949), many works have dealt with the analysis of the city size distribution of several groups of cities worldwide, generally finding an agreement of a power law distribution with observed data (e.g., Moura and Ribeiro 2006; Gligor and Gligor 2008; Giesen and Südekum 2011; Rastvorstseva and Manaeva 2016). A good review of plus and minus points identified in analyzing the so-called Zipf's law is given in Arshad, Hu, and Ashraf (2018).

Nowadays, it is also known that power law distributions are common in complex systems whose pattern is governed by the interactions between the constituent elements rather than by their specific nature and this makes the study of these systems particularly inclined to a mechanical-statistical analysis (Bogacz, Burda, and Wacław 2006; Dorogovtsev, Mendes, and Samukhin 2003). Among these systems, geographical ones are included, for which a number of studies have made use of entropy models (e.g., Wilson 1967; Batty 1974, 1976, 2010). Such an

Correspondence: Mauro Fiorentino, Department of European and Mediterranean Cultures: Architecture, Environment and Cultural Heritage (Dicem), University of Basilicata, Matera, Italy.
e-mail: mauro.fiorentino@unibas.it

Submitted: December 24, 2021. Revised version accepted: September 14, 2022.

doi: 10.1111/gean.12348

approach has not yet been sufficiently explored to analyze the city-size probability distribution, which certainly depends on the complexity inherent in the self-organization process of a group of cities, with particular reference to their size.

With regard to theoretical aspects, our article follows this furrow, as it exploits what was proposed by Sanchirico and Fiorentino (2008), who invoked the principle of maximum entropy (Jaynes 1957), which states that the least biased probability distribution of an unconstrained variable is the uniform distribution. By applying this concept to a complex system made up of many elements interacting between each other, they showed that the probability of an element to interact with other elements comes out to be a power law because the natural tendency to uniformity of this probability is contrasted by the tendency exhibited by the probability distribution of the interactions themselves to be the most uniform with respect to a homogeneous class, namely identified by the number of the connected elements. This contrast leads to a competition between the information entropies (Shannon 1948) of these two distributions, which provides the power law distribution as the most probable one for the number of interactions of a single element of the system.

A first objective of this article is mainly theoretical, consisting in using a statistical mechanics approach to derive the city size probability distribution, in order to show that this distribution arises as the most probable one when connections between cities become significant for the self-organization of a set of cities with regard to their dimensions. Still from a theoretical point of view, besides the static behavior of the probability city-size distributions, a rationale on how and why these distributions evolve in time is proposed in this study. This issue is here treated by proposing a simple model for describing the evolution of the city-size distribution for a given group of cities. By remaining in the framework of statistical mechanics, and using a thermodynamic analogy which allows us to give energetic interpretation to the distribution parameters, this model is used to discuss peculiar aspects shown by real-world city size distributions at different scales.

In this research, an empirical analysis is also carried out to support and validate the proposed theoretical developments, and to further speculate on their geographical and social implications. As a first step, population data of the European nations were used to support the crucial idea that the probability distribution of population sizes cannot be a power law if the administrative entities considered are not strongly integrated institutionally (common rules, common culture, common language, etc.) and economically (comparable needs, social services, industrial vocation, etc.), and then those power laws can fit to city-size probability distributions but not to the nation populations ones. This is in agreement with what stated by Cristelli, Batty, and Pietronero (2012), who observed that to achieve power laws cities should be homogeneous, in the sense that they should be integrated institutionally and economically.

Then, census population data of about 8,000 (7,914) Italian towns were used with the aim of fitting and interpreting the city-size probability distribution. Analyses were developed on a national basis as well as at a regional scale. For all groups of towns considered, a break point in the power law distribution of city sizes was identified. Consequently, the power law regarding the probability distribution for largest towns in each group was estimated. In this article, the role played by the break point in dividing the sample of cities considered into two groups of cities was particularly considered because it indicates, for instance, that these groups may not be homogeneous for such characters as for instance job opportunities, education services, health assistance, etc.

To summarize, the main objectives of these empirical analyses were: (i) validating the theoretical approach proposed to account for the emergence of a power law for the city-size probability distributions, and more in general supporting all the theoretical parts of the article; (ii) exploiting a thermodynamical analogy to interpret the power law parameters as crucial quantities, suitable to discriminate between different groups with regard to their energetic content; (iii) demonstrating that on a national basis the cities are better socially and economically integrated if their dimension is above a given threshold, which for the Italian system is equal to 10,000 citizens; (iv) showing that smaller towns, which are excluded by this group, are anyway linked to closer bigger towns on a regional basis, where a sufficiently large city behaves as a local hub. Yet, even at the regional scale a breakpoint is recognized, indicating that in all regions there are very small towns that are not integrated, neither at the national scale nor at the regional one.

Finally, it may help remarking that the article contains some mathematical developments that for the sake of lightness are reported in the Appendix. The reader more interested is referred to the Appendix for the full comprehension of the theoretical approach proposed here.

Theory

The probability distributions for describing city sizes

A set of cities in a country can be thought of as a thermodynamic system, subject to the energy exchange with the environment while the collection of interacting elements moves toward an equilibrium configuration. In line with this idea, in this section, we develop a statistical-mechanical formalism to derive the city size distribution.

In order to explain our method, let us start from considering a country with a total population of N people that live in a system of A cities, and denote by n_i the number of inhabitants of the ith city. The probability that a person chosen at random lives in a particular city i is then given by

$$q_i = \frac{n_i}{N}, \tag{1}$$

with

$$N = \sum_i n_i, \tag{2}$$

for index i ranging from 1 to A.

For our purposes, rather than referring to the distribution (1), it is worthwhile to group cities by the number of inhabitants. Let a_n be the number of cities with n citizens, the probability that a randomly selected city has n inhabitants is then given by

$$p_n = \frac{a_n}{A}, \tag{3}$$

with

$$A = \sum_{n=n_0}^{n_M} a_n, \tag{4}$$

where n_0 and n_M denote the less populous town and the city with the largest number of inhabitants, respectively. In this article, for the sake of simplicity, the sum in (4) is understood in the limit for large n, while n_0 is assumed to be equal to 1, and the role played by this parameter will be specifically discussed later.

In addition, the product na_n gives the total number of people that live in the cities with n residents, so that the frequency

$$r_n = \frac{na_n}{N}, \tag{5}$$

gives the probability that a person chosen at random lives in a city of n inhabitants. Note that the total number of citizens living in the whole system is also given by $N = \sum_n na_n$, where the sum is performed over classes of population n, whereas in equation (2) the constant N is summed over the set of all individual cities. By consequence, compared to the primary distribution (1), the r_n probability stems from an aggregation of people over a number of homogeneous cities with respect to population, once defined the probability p_n that supplies the frequency of cities with the same population n. However, as we will see in a moment, the information amount of the detailed distribution (1) is the same as the one contained into the aggregate distributions p_n and r_n. Finally, note that combining equations (3) and (5) provides

$$r_n = \frac{np_n}{\langle n \rangle}, \tag{6}$$

where

$$\langle n \rangle = \sum_n np_n = \frac{N}{A}, \tag{7}$$

denotes the average city size.

Following the information theory, the constituents of a system of cities stand for the particles of an equilibrium thermodynamic system, whose entropy acts as a measure of the average uncertainty encoded by the probability distribution of the elements. The Shannon entropy (Shannon 1948) may thus be defined for both p_n and r_n distributions as follows:

$$S_p = -\sum_n p_n \ln p_n, \tag{8}$$

and

$$S_r = -\sum_n r_n \ln r_n. \tag{9}$$

The entropies S_p and S_r quantify the randomness of the probability distributions p_n and r_n with respect to which a city in the first case and an inhabitant in the latter belong to the population class n.

It is worth noticing that either entropy function (8) or (9) is enclosed into the Shannon entropy

$$S_q = -\sum_i q_i \ln q_i, \tag{10}$$

defined for the distribution (1). This entropy measures the randomness degree about the location of an inhabitant within the system of cities, according to the knowledge of the special city which an inhabitant lives in. As explained in Appendix A.1, up to an additive constant, it is straightforward to prove that

$$S_q = -\sum_n r_n \ln \frac{r_n}{p_n}, \tag{11}$$

which makes clear as to what is stated about the role played by the p_n and r_n distributions in describing the population spreading over a set of cities. Equation (11) indeed states the average information at a finer scale, related to the q_i distribution, can entirely be deduced by the coupled distributions p_n and r_n, provided the r_n distribution be conditioned to the knowledge of the city weights p_n.

The Lagrange multiplier method to derive the city size distribution

Now let us pass to explain the emergence of a power law in general and Zipf's law in particular as a consequence of an equilibrium state. To do this, we need to review a few aspects of the interplay between statistical mechanics and information theory, which is at the heart of the Jaynes principle of maximum entropy (Jaynes 1957). Exactly as the particles of a thermodynamic system reach the most probable configuration at the equilibrium state, the least biased distribution of a random variable is the one that maximizes the Shannon entropy, subject to a set of constraints expressed as ensemble averages. Under conditions imposed by both the system structure and the energy exchange with the environment, the maximum entropy distribution is thus the most random, where random is to be meant as uniform (see Appendix A.2).

Following the theoretical framework originally developed by Sanchirico and Fiorentino (2008), and in order to adapt the approach to a geographical setting, let us assume for the moment that the set of cities, which we denote by G_A, be an isolated thermodynamical system. This means that no conditions except the normalization one affect the tendency towards the most probable state of the city size distribution. As explained in Appendix A.2, the unconstrained maximum entropy principle leads to the uniform distribution $p_n = 1/\Omega$, which implies that each of the Ω clusters of G_A has the same number of elements. The city size distribution is thus the most random. An isolated system of cities indeed would organize itself irrespectively of the relation structure of the inhabitants that compose it, which therefore plays no role in arranging cities into clusters of G_A. Similarly, considering the set of citizens, which we note G_N, as isolated from the rest of the universe, the most probable distribution of the citizens by population classes is the uniform distribution $r_n = 1/\Omega$. Thus, without any a priori bias about the number of cities that aggregate them, citizens would tend to fall inside available classes as randomly as possible. Under conditions of maximum randomness, cities and citizens then would tend to satisfy the uniform distributions $p_n = 1/\Omega$ and $r_n = 1/\Omega$, respectively, but this is not consistent with observations of real country distributions. This is because the statistical distributions of the cities and the citizens are related to each other by means of equation (6), which expresses the fact the two sets are not isolated. By consequence, any attempt by the elements of G_A to assume the configuration that satisfies the uniform distribution of the cities will unbalance the elements of G_N, whose corresponding distribution will move away from the uniform one, and vice versa. To be more precise, the particular city size distribution $p_n = 1/n\Lambda_2$, which unbalances the unbiased distribution $p_n = 1/\Omega$ with Λ_2 being a normalization constant, corresponds to the uniform probability $r_n = 1/\Omega$ via equation (6). From a geographical perspective, the uniform distribution $p_n = 1/\Omega$ comes from the fact that the set of cities is assumed to be impermeable to any energy exchanges from the surroundings, included the ones produced by the actions of the constituents of G_A itself. On the contrary, the enormous and diversified number of socioeconomic relations among citizens actually immerge the set G_A in a heat bath, wherein the people interactions play the role of an energy exchange with the environment. It is thus clear that, while driving its own distribution by size classes, the inhabitants affect the city size distribution too. By consequence, we may suppose the coupled sets G_A and G_N will both tend to an intermediate equilibrium state at which the attempt to get uniform the respective distributions p_n and r_n mutually balance. In this way, in the absence of other additional constraints, the competition that emerges between the sets of cities and citizens represents the least biased information by which the statistical state of the system is described. This phenomenon, which we will refer to as the *entropy competition*, can be taken into account by imposing the condition of maximum randomness of the elements of G_N, as a constraint for maximizing the entropy (8) associated to the set G_A.

As discussed in Appendix A.3, the problem is solved by searching for the maximum of S_p with respect to the following constraints:

$$\sum_n p_n = 1 \text{ and } \sum_n p_n \ln n = \langle \ln n \rangle_p, \tag{12}$$

with the latter expressing the entropy competition outlined above, so that the following pair of power-laws is achieved:

$$p_n = \frac{n^{-\gamma}}{\zeta(\gamma)} \text{ and } r_n = \frac{n^{1-\gamma}}{\zeta(\gamma - 1)}, \tag{13}$$

where $\zeta(\gamma) = \sum_{n=1}^{\infty} n^{-\gamma}$ is the Riemann zeta function that converges for $\gamma > 1$.

Yet, when analyzing real-world data, one can observe that few city size distributions follow a power law over the entire range of population values. On the contrary, below a threshold value, say n_0 much greater than 1, the city size distributions often deviate from the power-law regime. Thus, when plotting data on a logarithmic scale the straight-line confirms itself only for population values greater than n_0. It follows that the right-hand tails of the probability distributions fitting the real-world population values are given by

$$p_n = \frac{n^{-\gamma}}{\zeta\left(\gamma, n_0\right)} \text{ and } r_n = \frac{n^{1-\gamma}}{\zeta\left(\gamma - 1, n_0\right)}, \tag{14}$$

where the normalization factor $\zeta\left(\gamma, n_0\right) = \sum_{n=n_0}^{\infty} n^{-\gamma}$ is the incomplete zeta function (Newman 2005). In the next section we will discuss the geographical implications due to the threshold value n_0.

Before closing this section, it is useful to discuss some technical details. In the section devoted to the empirical analysis, in order to validate our model, rather than plotting a histogram of the city population by binning values we will make a plot of the exceedance probability P_n, also known as the complementary cumulative distribution, which gives the probability of finding a city with population greater than n. The exceedance probability can easily be computed through continuous approximation, by assuming that n is a real variable, which allows us to deal with integrals in place of the harder special functions needed for the discrete case. The normalization factors in the denominators of equations (14) then can be rewritten as:

$$Z_p = \frac{n_0^{1-\gamma}}{\gamma - 1} \text{ and } Z_r = \frac{n_0^{2-\gamma}}{\gamma - 2}, \tag{15}$$

from which we achieve the following probability density functions

$$p_n = \frac{\gamma - 1}{n_0}\left(\frac{n}{n_0}\right)^{-\gamma} \text{ and } r_n = \frac{\gamma - 2}{n_0}\left(\frac{n}{n_0}\right)^{1-\gamma}, \tag{16}$$

which replace the discrete distributions (14), respectively. A probability density function of such a type is also known as the Pareto distribution. Anyway, the city size distributions given by the first of equations (13), (14), and (16) differ only for the normalization conditions, while a power-law is obeyed in all three long-tails with the same scaling exponent γ. Finally, by definition of exceedance probability, we get:

$$P_n = \left(\frac{n}{n_0}\right)^{1-\gamma} \sim n^{1-\gamma} \text{ and } R_n = \left(\frac{n}{n_0}\right)^{2-\gamma} \sim n^{2-\gamma}, \tag{17}$$

where R_n gives the probability that a citizen chosen at random lives in a town of size greater n. In equations (17) we also made explicit the asymptotic power-law tails of the cumulative distributions for $n > n_0$. Finally, it is worth emphasizing that the exceedance probability P_n is actually Zipf's law, except the coordinate axes are flipped. More precisely, the power-law exponent of the city size distribution is related to Zipf's exponent by the relation $\gamma = 1 + 1/z$, from which one deduces that γ is equal to 2 for a perfect Zipf law ($z = 1$).

A rationale for analyzing evolution of the city size distribution

In this section, we introduce a two-stage model which aids identification of conditions about the population dynamics under which one can expect the entropy competition to emerge as a driving phenomenon for the system of cities.

In the previous subsection we introduced the distribution (1) giving the probability that a person chosen at random lives in a particular city i. Let us now denote by Π_i the probability that a person in a country decides to go and live in a city i. This latter probability, which is in principle different from (1), could simulate the internal migration at a national scale, when the person is randomly selected from a fixed set of citizens, or could mimic the demographic growth when a new citizen is injected into the system of cities. In general, Π_i will depend on a large number of variables characterizing the attractiveness of city i. We make here the first-order hypothesis that at a complexity stage the probability Π_i be proportional to the number of inhabitants of the city i, which is taken as a surrogate for the attractive strength of the city. It is indeed reasonable to assume that the more the size of a town, the more is the net of socioeconomic relationships that their inhabitants or business companies establish with the other cities, so as to attract more and more people from all over. It follows that there is a higher probability that a person goes to live in a city that already has a large number of inhabitants. Under the above assumption, the probability Π_i is just equal to the distribution q_i defined by (1). It is worth noticing that, from a geographical viewpoint, this hypothesis implies that citizens are associated to cities at a national scale, since people select the town in which spending their life by size, independently of how far is the city they choose from that of birth. Here we make clear the relevance of the probability Π_i to the entropy competition outlined above. To do this, let us deal with the population classes rather than individual cities. Then, the probability Π_n that a person chooses to live in any of the n-size cities is equal to the r_n distribution given by equation (5). We can simply refer to the aggregated probability Π_n rather than the detailed Π_i, since, by our hypothesis, it does not matter which particular city of n inhabitants a person chooses as all cities of n size are indistinguishable from one another as regards attractiveness.

Now let us focus our attention on the precomplexity phase. At an ancient stage of the system evolution, due to a number of features that can be traced back to a strong cultural inhomogeneity and high economic differences between geographical areas where cities develop, we can hypothesize that the weak interactions between inhabitants of different regions cannot exert any control on the system organization. Before complexity emerges, the probability that a person chooses to live in a given town thus depends on local factors only, which can be traced back to a native attracting force rather than to the city sizes. This means that citizens are associated to cities at a local scale, and people predominantly stay in the town where they were born. Since no one chooses his city of birth, the probability that a person ends up to live in a city i can be assumed to be uniformly distributed at random, according to the frequency $\Pi_i = 1/A$, which also maximizes the entropy (10) without any constraints except the normalization one. Note that this uniform distribution implies that each city has the same number of inhabitants, that is, $n_i = N/A = \langle n \rangle$, for all i from 1 to A, where $\langle n \rangle$ is the mean value of the distribution p_n, according to equation (7). Moving from cities to population classes again, we see immediately that the aggregated distribution Π_n, corresponding to $\Pi_i = 1/A$, is given by Kronecker's delta

$$\Pi_n = r_n = \delta_{n\langle n \rangle}. \tag{18}$$

This is a delayed discrete unit impulse that is zero everywhere except at $n = \langle n \rangle$, where it assumes a value of 1. There exists indeed a sole population class given by the average city size at which all the system population is concentrated. Two important consequences can be drawn from this result. The first consideration is that the distribution (18) is not useful for any statistical inference, since it represents the probability of an event that occurs with certainty (the probability of having the population class $\langle k \rangle$ is 1). Therefore, the average information is null, that is, $S_r = 0$ for $r_n = \delta_{n\langle n \rangle}$. It follows no entropy competition arises at this phase as no inference problem can be set for the unit impulse distribution. Secondly, a characteristic population scale emerges, which gives the only city size value that has a nonzero probability of occurrence. By consequence, the mean city size $\langle k \rangle$ naturally appears as the only average condition driving the system organization in the precomplexity phase, which is thus controlled by only two parameters: the number of cities and citizens. The average size of the cities then represents the only constraint reflecting the uniform distribution $\Pi_i = 1/A$ under which to search for the maximum value of entropy S_p. As shown in Appendix A.2, the least biased distribution achieved by maximizing the entropy (8) subject to the constraint (7), as well as the normalization one, is the Boltzmann exponential one:

$$p_n = \frac{e^{-\beta n}}{\sum_n e^{-\beta n}}, \tag{19}$$

where the normalization factor at denominator converges for $\beta > 0$. By assuming that n be a continuous variable so as to exchange the summation sign by an integral, the normalization factor rewrites as

$$Z(\beta) = \sum_{n=1}^{\infty} e^{-\beta n} \sim \int_0^{\infty} e^{-\beta n} dn = \frac{1}{\beta}. \tag{20}$$

It follows that the probability density function corresponding to the discrete one is given by $p_n = \beta e^{-\beta n}$, from which the exceedance probability is

$$P_n = e^{-\beta n}. \tag{21}$$

The thermodynamic analogy

Now, a couple of questions may arise: what does all this mean from an energetic viewpoint? And how the energies controlling the system before complexity emerges are transformed during the evolution process? In order to accommodate these issues, let us inspect the power-laws of equations (13) in the light of a thermodynamic analogy. The method outlined in the previous section leads indeed to a thermodynamic description of the system of cities that allows us to give an energy meaning to the constraints we used for the entropy maximization. Since this is a crucial point for a physically founded interpretation of our approach, we have reserved for this subject the Appendix A, which the reader is deferred to for a more comprehensive analysis. Here, it suffices to recall that for an equilibrium system at a temperature $T_p = 1/\gamma$, the distribution of energies over all accessible levels is described by the partition function, which for a complex system of cities is given by the normalization constant for the power-law city size distribution:

$$Z_p = \zeta(\gamma) = \sum_{n=1}^{\infty} n^{-\gamma} = \frac{1}{\gamma - 1}, \tag{22}$$

where the sum over population n is understood to be a sum over all the energy states the cities can assume, while the last equality holds in the limit for continuous approximation. Note that using the normalization factor at denominator of the first of (13) is the same as setting $n_0 = 1$ in the first of (15). The role exerted by the threshold parameter will be discussed later. From the knowledge of the partition function, all other thermodynamic variables can be derived. In particular, in Appendix A.3, it is shown that the partition function plays the role of generating function for the internal energy E_p, according to:

$$E_p = -\frac{\partial \ln Z_p}{\partial \gamma} = \langle \ln n \rangle_p = \frac{1}{\gamma - 1}, \tag{23}$$

where the last equality descends from the continuous expression for Z_p given by (22). Importantly, equation (23) states that the logarithmic mean of the population, that is, the second constraint of equations (12), acts as the internal energy of the system of cities. Analogously, with regard to r_n, the partition function for a system of citizens in a thermal equilibrium at a temperature $T_r = 1/(\gamma - 1)$ is given by the normalization factor:

$$Z_r = \zeta(\gamma - 1) = \sum_{n=1}^{\infty} n^{1-\gamma} = \frac{1}{\gamma - 2}, \tag{24}$$

in agreement with the second of (13), while the last equality is to be meant to hold in the limit for the continuous approximation, again. As above, the internal energy of the system of citizens is then defined by

$$E_r = -\frac{\partial \ln Z_r}{\partial \gamma} = \langle \ln n \rangle_r = \frac{1}{\gamma - 2}, \tag{25}$$

where subscript refers to the fact the logarithmic mean is calculated on the basis of the r_n distribution. Moreover, the Helmholtz free energies for a complex system of cities and citizens can be defined up to the Boltzmann constant as:

$$F_p = -T_p \ln Z_p = \frac{1}{\gamma} \ln(\gamma - 1) \text{ and } F_r = -T_r \ln Z_r = \frac{1}{\gamma - 1} \ln(\gamma - 2), \tag{26}$$

where we made use of equations (22) and (24), as well as the definitions for temperatures T_p and T_r.

On the other hand, as explained in Appendix A.2, for a precomplex system of cities in a thermal equilibrium at a temperature $T = 1/\beta$, the role of internal energy E is played by the average population $\langle n \rangle$ according to

$$E = -\frac{\partial \ln Z(\beta)}{\partial \beta} = \langle n \rangle = \frac{1}{\beta}, \tag{27}$$

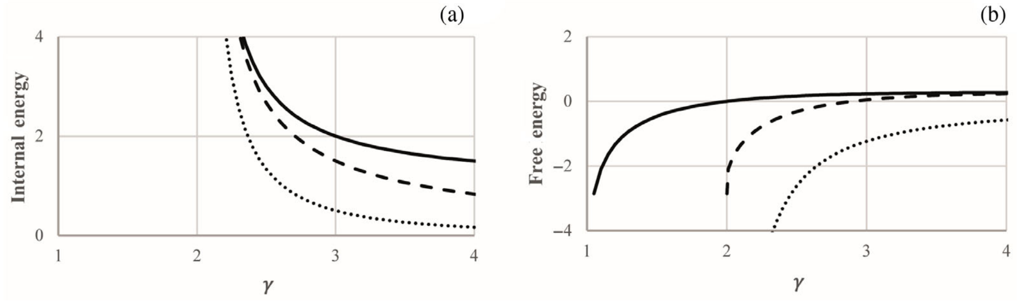

Figure 1. (a) Internal energy associated to a power-law distribution with scale parameter γ; solid line: $E_r - E_p$, where E_p and E_r are given by equations (23) and (25), respectively; dashed line: $E_p + E_r$; dotted line: Internal energy associated to the exponential distribution (equation 27), where parameter β is estimated by means of the relation (30). (b) Free energy as a function of the scale parameter γ; solid line: Free energy F_p associated to the power-law distribution p_n (first of equation 26); dashed line: Free energy F_r associated to the power-law distribution r_n (second of equation 26); dotted line: Free energy associated to the precomplexity exponential distribution, according to equation (28), where β is given by equation (30).

where the mean is performed with respect to the exponential distribution (19), while the last equality comes from using (20). Finally, the free energy is given as a function of parameter β as follows:

$$F = -T \ln Z(\beta) = \frac{1}{\beta} \ln \beta. \tag{28}$$

The dependence of the internal and free energies on the scaling exponent γ is shown in Fig. 1, where parameter β has been estimated by assuming the same average population value for both the exponential distribution (19) and the power-law city size distribution (first of (13)), this latter being given by

$$\langle n \rangle_p = \frac{\gamma - 1}{\gamma - 2}. \tag{29}$$

By comparing equations (27) and (29), one deduces that

$$\beta = \frac{\gamma - 2}{\gamma - 1}. \tag{30}$$

These figures suggest the following comments. Firstly, as the condition that ensures the process be exergonic with a flow of energy from the system to the surroundings is that free energy decreases, Fig. 1b shows that in the range of values of γ observed for real systems ($\gamma < 4$) evolution occurs by changing this parameter from higher to smaller. Also, this flow direction is clarified in Fig. 1a, which demonstrates that the internal energies for both p_n and r_n come from the initial energy of the system, being $E_p + E_r$ less than $\langle n \rangle$, which represents the only internal energy of the system before complexity emerges. In addition, Fig. 1a shows that once the average information related to the distribution r_n becomes significant, the internal energy previously owned by the no-complex system is divided between both the ensembles of cities and citizens; in particular, one can see that until γ is greater than 3 the internal energy is more or less equally shared between p_n and r_n, since $E_r - E_p \sim 0$. Furthermore, in the same range, evolution does not require a high increase of total energy. On the contrary, while γ approaches

2, E_r becomes greater and greater than E_p and, in the limit for $\gamma \to 2$, E_r diverges, whereas E_p attests itself to a finite value, according to equations (23) and (25). Therefore, for lower values of γ, much more energy is needed by the system to move toward $\gamma = 2$ or, in other words, a great amount of new population is necessary, which is acquired as internal energy by r_n. This means that in this range the city-size distribution is strongly controlled by the probability r_n and the residual local attractiveness progressively loses its strength as new energy is available. Thus, it is suggested that Zip's law ($\gamma = 2$) represents the dominance of the system of cities as a whole on local features, which in our schematization may occur when the internal energy of the system of citizens diverges.

Also, it is noteworthy that all curves in Fig. 1a show a knee-shaped behavior for γ below 3, so that the system of cities may easily reduce γ to 2.5 but faces much more difficulties to go below this value. In other words, local attractiveness, although weak in the competition with the system power, is unlikely to completely disappear. One more consideration is reserved for the behavior of free energy, which shows differences between the three distributions taken into account here. As shown in Fig. 1b, F_p is always positive for $\gamma > 2$, whereas the free energy is negative for the exponential distribution, thus pointing out that in each town the population may exert work on the surrounding system, while this is obviously false when the ensemble of cities does not constitute a system because there is no relation between its elements. Also, it is interesting to note that F_p tends to zero when γ approaches 2.

In conclusion of this section let us briefly comment on the role played by the threshold population value n_0 that appears in equations (14). As discussed in the previous section, this is indeed the lowest population value at which the power law is obeyed. As we will prove in our empirical analysis, when inspecting the entire ensemble of cities one may note that the exceedance probability of the size n tends indeed to vary very slowly below a threshold n_0, which is thus identified as a cutoff point in the city size distribution. An interpretation of n_0 can be suggested by the evolution model proposed here, by pointing out that this value represents the minimum city size taken into account by people to choose the city for living by moving from other towns. Yet, p_n becomes a power law when the average information about the citizen distribution by population classes prevails on the mean city size and this occurs when mobility between towns becomes a significant practice. Under this assumption, the tendency to the uniformity of r_n represents indeed a constraint for p_n to evolve far from the exponential distribution. Below n_0, the town population amount is controlled by local factors only. In addition, this cutoff value is expected to decrease during the evolution process. In fact, in the beginning, just a few very large cities become attractive beyond their surroundings, so that the population classes related to the citizen set G_N deal with large values of n only, and $n = n_0$ is sufficiently high too. At this stage, cities with population lower than n_0 do not participate to mix people between towns, unless suffering the emigration process activated by the increasing attractiveness of largest cities. Afterwards and progressively, more towns of less high size start to attract new inhabitants themselves, and $n = n_0$ decreases.

Empirical analysis and discussion

In this section data analysis to support and validate the theoretical developments outlined above and to shed more light on their geographical and social implications is carried out, with particular regard to the way a system of cities tends to organize itself. Trying to respect a logical track, the results are presented in the following order.

First, the hypothesis that the city-size probability distribution in the initial state – in which the ensemble of cities does not represent an interconnected complex system – can confidently be assumed as an exponential distribution is assessed. Then, paying more attention to the core of the article, which is mainly focused on the city size distributions nowadays observed, the analysis is aimed to: (i) demonstrating how good is the conflict of entropy model to interpret these distributions; (ii) to discuss the presence of a *cut-off* value in these distributions and its meaning, even in the light of the proposed evolution model; (iii) showing that on a national basis, the cities are better socially and economically integrated if their dimension is above a given threshold; iv. demonstrating that, even if at the regional scale a breakpoint is recognized also, smaller towns are anyway linked to closer bigger towns, where a sufficiently large city behaves as a local hub. As regards the first issue, since it deals with a very ancient state of the city size distribution, there is no way to confirm our hypothesis on a rigorous experimental basis. However, we may assume that a system of countries, which have evolved with negligible mutual interchange according to very different constraints (culture, language, etc.), may represent a useful analog of a group of cities at the pre-complexity stage. More precisely, for our purposes, we may state that there is similarity between the shape of a current country-size (population) distribution and a precomplexity-stage city-size probability distribution.

To this aim, we analyzed population data of 45 European countries as given for the year 2019 by Eurostat in the open data section *Population change - Demographic balance and crude rates at national level* (online data code: DEMO_GIND). These data were ordered to estimate their exceedance probabilities shown in Fig. 2, where a data fit by the exponential probability law of equation (21) is also provided (Fig. 2a).

It is worth noticing that the right-hand tail, which refers to larger countries where sociocultural features are more different from each other, and then are those more similar to cities at a precomplexity stage, shows a good agreement between data and our theoretical law (equation 21). Thus, even acknowledging the limits of the assumed analogy, the hypothesis that the exponential distribution can confidently be assumed to represent the initial city-size probability distribution is (indirectly) empirically supported. To further point out this statement and to enforce the validity of the hypothesis underlying the evolution model we proposed, in Fig. 2b it is shown how a power law is absolutely inadequate to fit the right tail of the analyzed empirical distribution, so confirming that the Pareto-type distributions are not suitable for not complex systems.

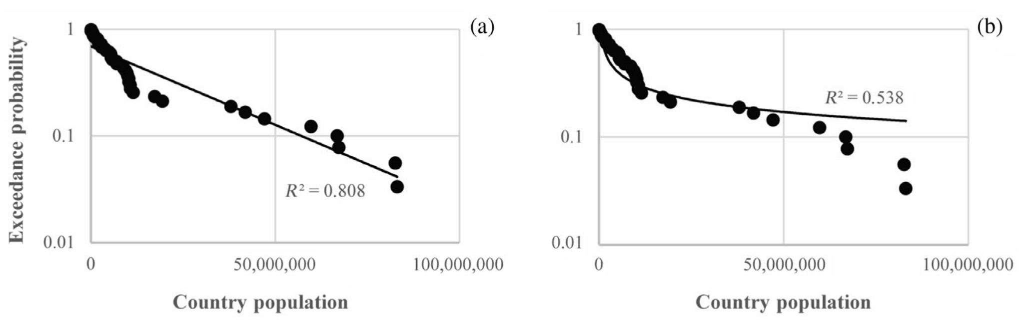

Figure 2. (a) Europe zone: Exceedance probability of country size (population amount) in a semi-log plot. Observed data (dots) drown according to Hazen plotting position, and fitted exponential probabilistic function (solid line) of equation (21); (b) log–log plot of the same data fitted by a power law.

Table 1. Italy: Administrative Units and Some Geographic Features

Regions		Number of provinces	Number of municipalities	Population			Surface (km^2)	% Mountain area
				Total	Min	Max		
1	Piemonte	8	1,181	4,328,565	32	857,910	25,387	43.3
2	Valle d'Aosta	1	74	125,653	85	33,916	3,261	100.0
3	Liguria	4	234	1,532,80	61	565,752	5,416	65.1
4	Lombardia	12	1507	10,010,833	30	1,406,242	23,864	40.4
5	Trentino-A.A.	2	291	1,074,034	141	120,641	13,606	100.0
6	Veneto	7	563	4,884,590	127	259,087	18,407	29.0
7	Friuli-V. Giulia	4	215	1,210,414	102	201,613	7,862	42.6
8	Emilia-Romagna	9	328	4,459,453	69	395,416	22,453	25.3
9	Toscana	10	273	3,701,343	391	366,927	22,987	25.1
10	Umbria	2	92	873,744	94	164,880	8,464	29.3
11	Marche	5	228	1,520,321	111	99,077	9,401	31.0
12	Lazio	5	378	5,773,076	69	2,808,293	17,232	26.1
13	Abruzzo	4	305	1,300,645	82	119,862	10,832	65.1
14	Molise	2	136	303,790	104	48,337	4,461	55.3
15	Campania	5	550	5,740,291	224	948,850	13,671	34.6
16	Puglia	6	257	3,975,528	163	315,284	19,541	1.5
17	Basilicata	2	131	558,587	229	66,393	10,073	46.9
18	Calabria	5	404	1,912,021	211	174,885	15,222	41.9
19	Sicilia	9	390	4,908,548	187	647,422	25,832	24.5
20	Sardegna	8	377	1,622,257	77	151,005	24,100	13.6

To deal with the other objectives of the empirical analysis, let us pass to consider city's population data. In such a case, we used Italian population values of the 7,914 municipalities as certified by the Italian Institute of Statistics (ISTAT) at the end of the year 2019. The data vary from a very small town of 30 residents in the case of Monterone (province of Lecco, in the North), to the maximum value of millions of citizens, equal to 2,808,293 residents in the case of Rome, the largest municipality. The average population size is 7,536 with a standard deviation of 42.40, while the total population is 59,816,673 and the covered area is 302,073 km^2 (ISTAT 2019). To explore the relationships between the sets of citizens and their subgroups, the municipalities were grouped into 20 regions and 110 provinces, namely the higher-level administrative units, as shown in Table 1, where some territorial peculiarities are reported too. In order to better interpret the results that we are about to show, it may be useful to premise some general features of the Italian territory. In Italy, as well as elsewhere, there are several elements that make the city a material and symbolic space of social relations and it is difficult to give it a univocal definition in terms of size, heterogeneity and form. However, for the sake of simplicity, the urban area is here considered to be that part of the territory falling within the municipal administrative limits, although one should acknowledge that, in such a manner, we are neglecting that the post-industrial cities grown in the last 50–60 years are rapidly changing in their administrative form, morphology and social structure (Arbia 2001). The phenomenon of urbanization outlined with the centralization of the population and the simultaneous depopulation of the countryside and small villages has generated intense short-, medium-, and long-range migratory flows that are movements toward the larger provinces, from the south toward the north of Italy and also toward other countries. Furthermore, the Italian territory, as a physical

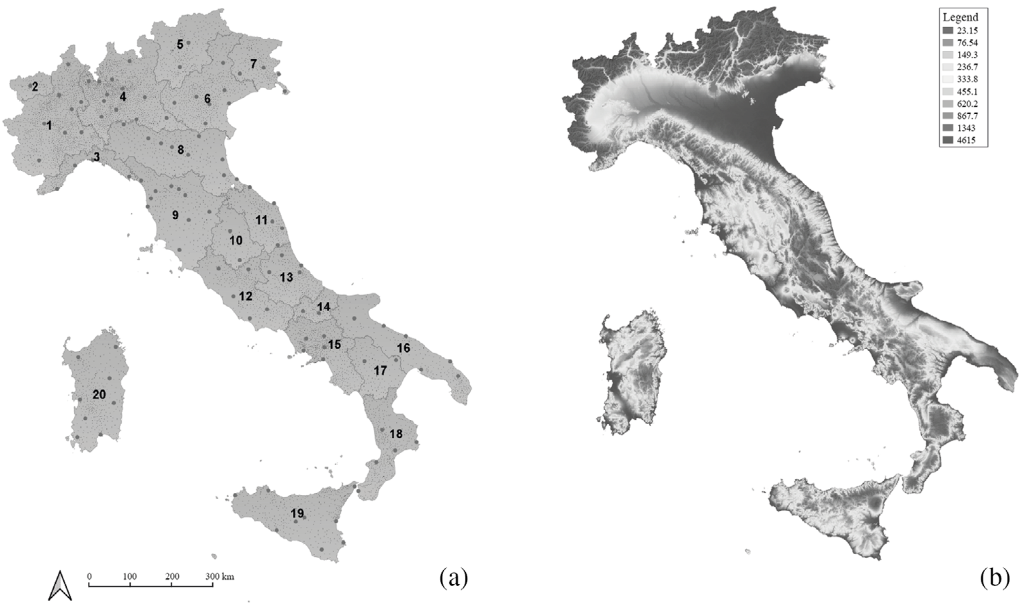

Figure 3. Italy: (a) map of territorial units (the numbers are referred to regions as in Table 1, bigger dots indicate provinces, smaller dots indicate municipalities); (b) elevation map.

place where social relations and production are woven, conditioning the quality of life and the economy, is characterized by very pronounced differences between the North and the South, between cities and the countryside and between lowlands and mountain areas, which largely overlap with inland areas. Today, as in the past, the position of territories and the resources available to them determine their development opportunities. Thus, a morphological description of the country enables us to identify the geographic criticalities that can contribute to hinder the socioeconomic development of a system of cities. The spatial pattern of all Italian territorial units is shown in Fig. 3a, while Fig. 3b shows the elevation map, which somehow reveals its impact on the evolution of the system of cities. The database described allowed us to verify the theoretical model and carry out the application reported below. In Fig. 4a, the log–log plot of observed exceedance frequencies of city size values for all Italian towns is shown with regard to the complementary cumulative probability distributions P_n and R_n. In the first case, frequencies were estimated by the Hazen plotting position (Hazen 1914) after simply ordering the population sizes of all censed municipalities, while in the latter the estimates of R_n were achieved for each city by summing the inhabitants of all the cities larger than the one in question, and by making the ratio of this sum to the total number of inhabitants in Italy. It is evident that the right-hand tails follow a power law in both cases for n greater than a threshold value that for the sake of simplicity can be assumed equal to 10,000. This behavior fully supports all the hypotheses we assumed in this work, with particular regard to the emergence of the entropy competition phenomenon described in the previous sections. This is well shown in Fig. 4b, where it is demonstrated that the power law provides a very good fit for both P_n and R_n, and that the difference between the estimated exponents of these laws is close to 1, as theoretically expected by invoking equations (17). With more specific regard to geographical aspects, these results indicate that people living in towns with a size greater than 10,000 belong to a statistically homogeneous group, where they may happen to live in a town rather than another according to

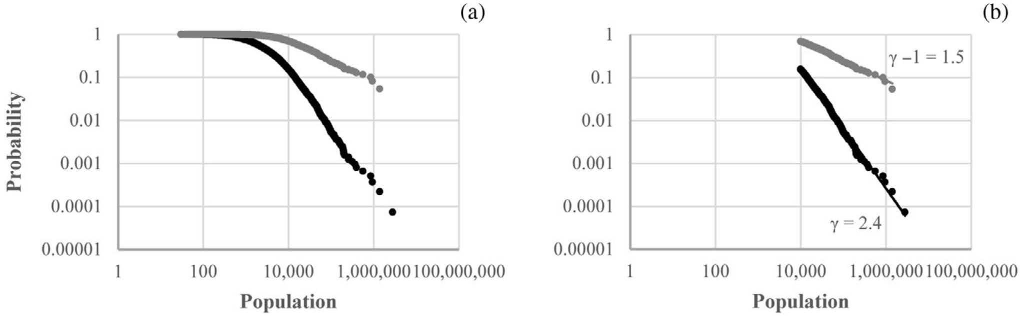

Figure 4. Exceedance probability of city size in Italy (year 2019 census). P_n (●) is the probability of a town to have a size greater than n, R_n (●) is the probability of a citizen to live in a town of a size greater than n. The symbols show the observed frequencies, while the dashed lines indicate the estimated probabilities: (a) all municipalities in Italy; (b) municipalities with more 10,000 residents.

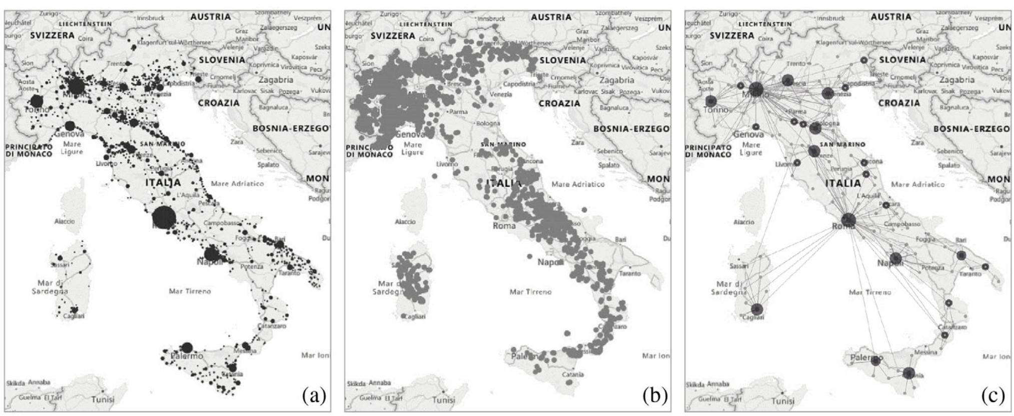

Figure 5. Italy. (a) Map of cities with more than 10,000 inhabitants; (b) map of cities with less than 1,000 inhabitants; (c) Italian mobility connection map referred to January 2020, third week.

the number of accidents and opportunities, but in situations that would not significantly change whether living in any other town within the group. This shows that these cities belong to the same complex system and take on similar social, economic, and relational behaviors. On the contrary, the towns with population smaller than 10,000, appear to be less connected to others and condemned to suffer for the lack of services accessible to people living in larger cities only. The maps shown in Fig. 5 offer an interesting sketch of how these two groups of towns are scattered within Italy. In particular, Fig. 5a shows that the cities with more than 10,000 residents are organized by a number of larger cities acting as hubs for groups of towns around them. It is evident that cities come in clusters where some bigger towns tend to aggregate some other smaller ones around them. As well as one can easily see that the system of cities self-developed following some preferential directions, which practically avoided mountainous areas (Alps and Apennines). It is also interesting to notice that conversely the smallest towns are mainly located inland and in mountain areas, constituting a system that tend to be geographically separated from the other (Fig. 5b). Comparison of Fig. 5a–c, whose similarity is evident, demonstrates

Table 2. City Size Distribution for Italian Regions: From left: Parameter Estimates (n_0 and γ) for the Power Law Fitted to Municipality Population Census Data for the Year 2019; Coefficient of Determination Achieved for $n > n_0$; ξ: Percentage of Population Living in Towns with $n > n_0$; ν: Percentage of Towns with $n > n_0$

Region	n_0	γ	R^2	ξ	ν
Abruzzo	2,500	2.12	0.971	84%	30%
Basilicata	2,500	2.38	0.977	84%	44%
Calabria	2,500	2.30	0.988	84%	42%
Campania	6,000	2.25	0.951	86%	34%
Emilia-Romagna	5,000	2.22	0.983	92%	59%
Friuli-Venezia Giulia	3,500	2.37	0.975	83%	36%
Lazio	2,000	1.96	0.944	97%	57%
Liguria	2,000	2.02	0.984	93%	42%
Lombardia	5,000	2.48	0.988	79%	31%
Marche	3,000	2.15	0.955	89%	46%
Molise	1,000	2.17	0.992	87%	48%
Piemonte	1,000	2.04	0.987	94%	50%
Puglia	10,000	2.45	0.990	83%	42%
Sardegna	2,000	2.16	0.992	88%	43%
Sicilia	7,000	2.25	0.982	86%	37%
Toscana	8,000	2.31	0.991	86%	42%
Trentino-Alto Adige	1,500	2.30	0.982	90%	55%
Umbria	2,500	1.96	0.966	94%	55%
Valle d'Aosta	500	2.19	0.926	94%	65%
Veneto	5,500	2.58	0.980	83%	46%

that the city pattern controls the number of mutual relations between towns, which in turn affect the system itself. The latter figure reproduces in a geo-spatial way the distribution of the weekly macro mobility flows among the Italian provinces and is based on the data of the City Analytics solution proposed by Enel X, aimed at producing statistical indicators to support the country during the first period of the Covid-19 emergency. In particular, Fig. 5c referring to the third week of January 2020, incidentally the period just before the outbreak of the pandemic in Italy, shows how much percentage of the population traveled between the provinces during the week considered (Enel X and Here - City Analytics 2020). In the figure that percentage is indicated by the thickness of the connections, although this is not significant for the present study.

The population data were analyzed at the regional level also, verifying similarities and dissimilarities with the national configuration. For each region, the city size probability distribution p_n was estimated, by identifying a power law above a breakpoint n_0, which was recognized in all cases as expected. Some results are shown in Table 2, where the estimated values of n_0, the estimates of the power law exponent γ, the coefficient of determination R^2 achieved by fitting the theoretical power law to data, the percentage of the population living in cities with a number of inhabitants greater than the breakpoint, and the percentage of cities above it, are reported. The n_0 values were estimated as those maximizing the R^2 for the power law referred to the city-size exceedance probability of the cities having population greater than n_0. Analyzing these

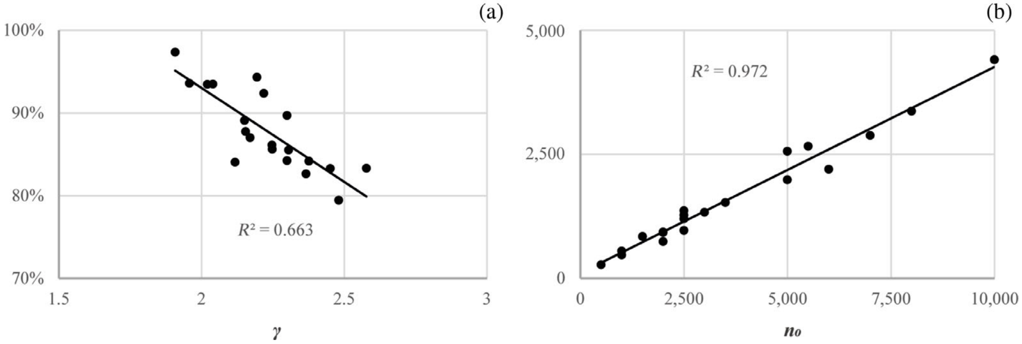

Figure 6. (a) Correlation between the percentage of population living in cities with $n > n_0$ and the estimates of γ for the right-hand tail of city size distributions in all Italian regions; (b) correlation between the average size (on the ordinate axis) of towns with $n < n_0$ and with n_0 for all Italian regions.

results, one may note the following significant issues. In almost all regions the breakpoint is much lower than 10,000 inhabitants as instead estimated at the national scale. This indicates that some smaller towns that do not participate in the more homogeneous group of cities at the national scale are somehow connected to a closer bigger city acting as a local hub. Consequently, one may think that these towns are likely to exploit the services provided nationally if and only if a local hub mediation is possible, and that a regional integration is found on the basis of acceptance of services and opportunities certainly lower than those fully provided at the national scale. In all regions, the percentage of population living in the group of towns above the breakpoint is always greater 80%, although the percentage of the number of towns above that point is less than 50% in the majority of cases. This indicates that although the percentage of inhabitants which are more likely excluded by the national system remains low enough, it is spread on a not negligible number of towns. As regards the orographic control, it is noteworthy that for all the regions where n_0 is at least 5,000 inhabitants the percentage of mountainous areas remains below 40%, thus indicating that the population self-organizes in not too small towns in sufficiently flat areas only. In Fig. 6a, it is shown that there is a correlation between the estimated values of γ and the percentage of population belonging to the group of larger and more integrated towns, with a tendency of the latter to decrease as the former increases. This is well interpreted in the light of the evolution model proposed in this article, suggesting that γ decreases toward the limit value 2 (Zipf's law) as more and more citizens are included into the complex system of larger cities. Also, this is in agreement with the second of equations (15). In Fig. 6b a very strong correlation between n_0 and the average size of towns having $n < n_0$ is shown, which indicates that the more the system of cities tends to organize in larger units the bigger are the smaller towns less connected to the remainder of the region. Although this is a somehow obvious result, the elevated value of the coefficient of determination ($R^2 = 0.972$) may suggest further future investigation. These results are in line with the meta-analysis carried out by Cottineau (2017), which related the scaling exponent γ to other geographical descriptors. Significant cases emerge for different regions, four of which are reported below. First, let us consider Lazio Region, where the largest city of the country is located. This is one of the regions for which the right-hand tail of the city size distribution closely approximates a perfect Zipf's law with $\gamma = 2$ (Table 2). Coherently a very low value of n_0 (2,000 inhabitants) and the highest

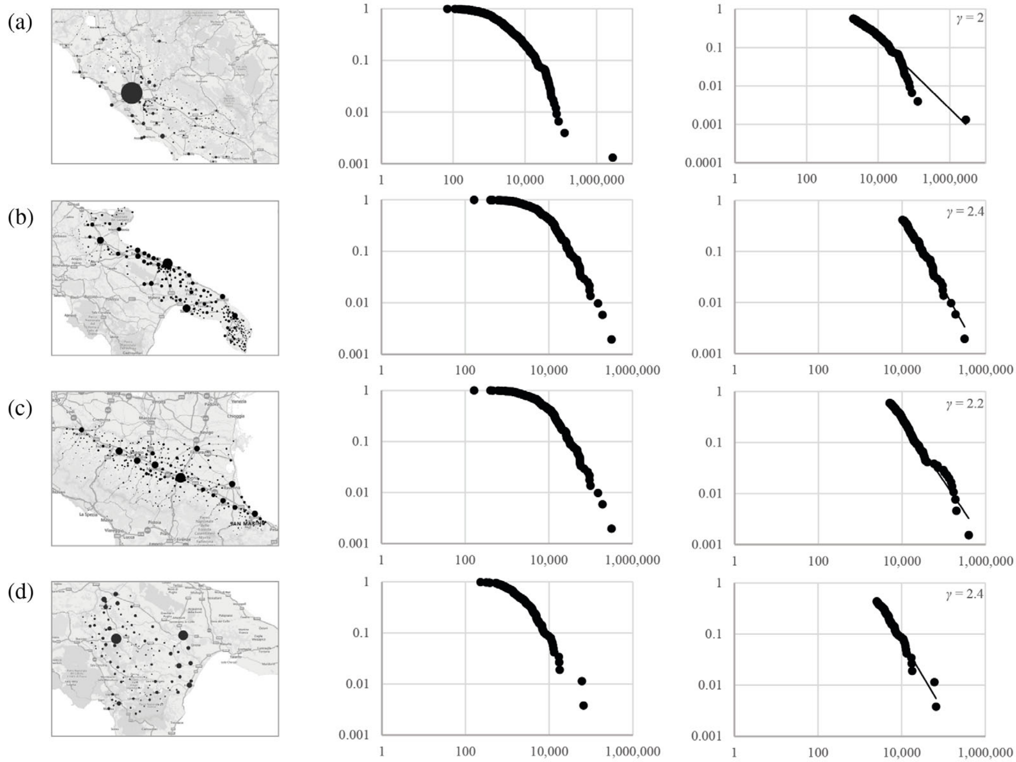

Figure 7. Four Italian regions: (a) Lazio, (b) Puglia, (c) Emilia Romagna, and (d) Basilicata. From left to right: City size sketch, exceedance frequencies for all the municipalities in the region, and exceedance frequencies for those municipalities which are well fitted by a power law distribution.

percentage of population living in towns with size greater than n_0 are observed too ($\xi = 97\%$). It seems evident of the influence of Rome as capital city, which acts as an attractive pole for the cities in its hinterland and redistributes the population more evenly along the coast, and in the less mountain internal areas (Fig. 7a). One should also notice a rather equal distribution of municipalities in the two ranges divided by n_0 ($\nu = 57\%$), meaning that there is a great number of very small towns, which will likely tend to aggregate with each other or disappear in the future. A different behavior is observed for the Puglia Region (South-East of peninsula), which is the only one amongst all that maintains a breakpoint n_0 equal to the one observed at the national scale (10,000 inhabitants). This is emblematic, as it demonstrates how favorable climatic and geographic features, such as flat orography and temperate climate due to the proximity to the coast, induce a higher connection level where population is more equally distributed, so that depopulation of territory is less significant (Fig. 7b). The case of Emilia-Romagna is reported here too, which in line with other regions in northern Italy, such as the Lombardy region, presents a level of homogeneity for city sizes greater than 5,000 inhabitants. In this case, it is evident how the municipalities are aggregated along the main – and historical – roads, so showing how important infrastructure networks are in the logic of settlement in the territory (Fig. 7c). Finally, let us consider the Basilicata Region (Fig. 7d), which is one the southern regions suffering for a still active depopulation process. Here, $n_0 = 2{,}500$ is almost low as in the Lazio Region but the city acting as a local hub is much smaller than Rome and the power law exponent of the city size distribution is still far from 2 (i.e., $\gamma = 2.4$). Thus, it is depicted as a region with a lower level of internal connection where its city size distribution can be seen as the mirror of a system that finds his homogeneity on the basis of acceptance of a lower standard for services and opportunities.

Conclusions

In this work attention was paid to the experimental evidence that right-hand tail of city-size probability distribution is mostly well fitted by a power law.

First, it shed more light on the way complexity affects this behavior by demonstrating that power law emerges as the maximum entropy distribution, once the complex connections between cities arise. In particular, it has been shown that the power law occurs when the city size organization is controlled by two probability distributions, the first providing the probability for a city to have a certain number or inhabitants and the latter giving the probability that a person resides in a town of a given number of citizens. It was also emphasized that all the towns contributing to that power law represent a group of elements which are integrated institutionally and economically, so that it cannot be excluded that a citizen moves from one city to another, regardless of city size. Second, a new two-stage model was hypothesized for the evolution of city-size systems, which was analyzed by exploiting the thermodynamical analogy that allows a physical interpretation of probability distribution parameters. In this light, it was demonstrated that a complex system of cities grows capturing towns progressively smaller into the city-size power law. Thus, the power law exponent γ continuously decreases and easily gets values below 3, while much more energy is needed to push it toward 2, that is, the value corresponding to well-known Zipf's law. All the assumed assumptions were strongly supported by the analyses carried out by using population data. European data were used to demonstrate that the right-hand tail of the nation-size probability distribution is fitted by an exponential law rather than a power law, indicating that the ensemble of nations is much less complex than that of the cities within each nation, and confirming the hypothesis that complexity underlying a power-law probability distribution is related to the linguistic and institutional homogeneity linking the elements of the set taken into account. The Italian population census data based on the year 2019 regarding about 8,000 Italian municipalities were used too. It was found that the Italian city system is complex and integrated for towns with more than 10,000 residents, for which the size distribution shows a good fit to a power law with the exponent $\gamma = 2.4$. In addition, it was demonstrated that the smaller towns, which do not contribute to the overall complexity, suffer an evident geographic and orographic control. Analyses were carried out at a regional level too, which showed that at this scale too systems of cities whose size distribution is well described by a power law are identified. In this case, different from the national scale, even towns smaller than 10,000 inhabitants are statistically linked to a larger city acting as a local hub. The minimum size of towns participating in the group leading to the power law is inversely correlated to the percentage of regional population living in this group of towns. The presence of towns that although disconnected at the national level keep a link to the less small cities in the region let us conclude that at the regional scale the complexity that induces homogeneity between towns is found on the basis that, in comparison to the national level, may tolerate the lack of services and opportunities, which are provided at this latter scale only. In conclusion, we would like to remark that this article reconciles city-size distribution analysis with the theory of complexity and may shed more light on the distribution inner structure too. In addition, the results we achieved indicate that the suggested hypotheses may confidently be assumed to drive new studies on the organization of ensemble of cities.

Acknowledgement

Open Access Funding provided by Universita degli Studi della Basilicata within the CRUI-CARE Agreement.

Appendix A

Shannon entropies

Let us derive equation (11), which expresses the local entropy (10) as a function of the two coarse-grained entropies (8) and (9). Moving from a sum over cities to a sum over population classes, it is easy to recognize that

$$S_q = -\sum_i \frac{n_i}{N} \ln \frac{n_i}{N} = -\sum_n a_n \frac{n}{N} \ln \frac{n}{N}. \tag{A1}$$

Equation (A1) then allows us to recast the entropy (9) in the following form:

$$S_r = -\sum_n a_n \frac{n}{N} \ln \frac{n}{N} - \sum_n r_n \ln a_n = S_q - \sum_n r_n \ln p_n - \sum_n r_n \ln A, \tag{A2}$$

where we used definitions (3) and (5). By making use of the normalization condition $\sum_n r_n = 1$, Equation (A2) immediately provides:

$$S_q = S_r + \sum_n r_n \ln p_n + \ln A = -\left(\sum_n r_n \ln r_n - \sum_n r_n \ln p_n \right) + \ln A,$$

from which

$$S_q = -\sum_n r_n \ln \frac{r_n}{p_n} + \ln A,$$

which only differs from relation (11) by constant $\ln A$.

Boltzmann statistics

Both Shannon entropy and Jaynes' principle can be founded on a solid statistical-mechanical basis. It can indeed be easily shown that equation (8) may be drawn from the Boltzmann entropy $S = K_B \ln W$, which counts the number W of microstates corresponding to a given energy configuration that the molecules of an ideal gas can occupy, where K_B denotes the Boltzmann constant. In our scheme, the gas of particles is represented by the statistical ensemble G_A of cities, whose elements are grouped in Ω clusters each of which containing a_n cities of size n. This allows us to interpret the population class n as the energy level of the nth cluster: as it is usual in thermodynamic statistics, the energy amount of the nth level can indeed be assumed to be equal to $E_n = ne$, so that every energy level is interspersed by a constant energy value equal to e. In this work, without losing generality, we will always assume $e = 1$, as we are interested in the dependence of the energy on the population value n only.

By simple counting arguments, the number W of ways of arranging A distinguishable cities in Ω clusters, with the nth cluster containing a_n cities of energy E_n, is given by the multinomial coefficient $W = A!/\prod_n a_n!$, for $1 \leq n \leq \Omega$. By substituting the expression for W in the Boltzmann entropy, and in the limit for large A, the Gibbs entropy of a system of cities is achieved, which is the same as the Shannon entropy (8) up to the Boltzmann constant. It follows the city size distribution p_n can be interpreted as the probability of finding a city in the nth energy level too. The same arguments can be used to derive the Shannon entropy (9), with the citizens in the place of cities. According to the above scheme, the Gibbs–Shannon entropy (8) also measures the disorder enclosed in the city size distribution due to the arrangement of cities in energy levels. The second law of thermodynamics thus imposes that, for an isolated system of towns, the most probable city size distribution is the one that maximizes the entropy (8), subject to the sole normalization condition:

$$\sum_n p_n = 1, \tag{A3}$$

which is nothing but equation (4) expressed in terms of probability p_n, where the sum ranges from 1 to the number Ω of energy levels. This distribution can be achieved by making use of the Lagrange multiplier method, which in this appendix is described by following Singh (2013).
In such a case, the Lagrange functional assumes the following expression:

$$\mathcal{L}\left(p_n\right) = -\sum_n p_n \ln p_n - \lambda \sum_n p_n + \lambda, \tag{A4}$$

where λ is the Lagrange multiplier corresponding to constraint (A3). By equating to zero the first derivative of (A4) with respect to p_n, one obtains the uniform distribution $p_n = 1/\Omega$, which states every energy level of the ensemble of cities is equally likely. Note that because of relation (6), the citizen distribution by population classes corresponding to $p_n = 1/\Omega$ is given by $r_n = n/\Lambda_1$, where the normalization factor Λ_1 is finite for a finite number of clusters. By proceeding as above, for an isolated system of citizens, the entropy (9) attains its unconstrained maximum value in correspondence of the uniform distribution $r_n = 1/\Omega$, whose related city size distribution via equation (6) is $p_n = 1/n\Lambda_2$, where constant Λ_2 is finite for finite Ω, again.
On the other hand, an isolated thermodynamic system is purely an ideal abstraction, since any system in a thermal equilibrium exchanges energy with the surroundings that act as a heat bath at a fixed temperature T. One thus expects some fluctuation of the internal energy around the time average value. Due to the ergodic hypothesis, the average over time can be assumed to be equal to the ensemble average of the energy over all levels accessible to the system. For a like-Boltzmann equilibrium configuration, which may realize at a precomplex stage, the internal energy of the system of cities is therefore given by the average population value:

$$E = \sum_n np_n = \langle n \rangle. \tag{A5}$$

Driven by the energy flow through thermal contact with an environment, a precomplex system of cities thus will eventually settle down to an equilibrium state at which the entropy (8) reaches the maximum value under the energy constraint (A5), plus the normalization condition (A3). In this case, the Lagrange functional assumes the following expression:

$$\mathcal{L}\left(p_n\right) = -\sum_n p_n \ln p_n - \lambda \sum_n p_n + \lambda - \beta \sum_n np_n + \beta\langle n \rangle, \tag{A6}$$

where β is the Lagrange multiplier associated to the constraint (A2.3). By making the first derivative of (A6) with respect to p_n equal to zero, we obtain the exponential distribution

$$p_n = \frac{e^{-\beta n}}{\sum_n e^{-\beta n}}, \tag{A7}$$

for the population n, according to the Boltzmann statistics for the energy configuration of a gas of particles. The Lagrange multiplier β thus plays the role of an inverse generalized temperature $T = 1/\beta$. Accordingly, in the limit for $\beta \to 0$ (high temperatures), (A7) boils down to the uniform distribution $p_n = 1/\Omega$, thus regaining the unconstrained equilibrium state for the ensemble of cities. Importantly, derived by using condition (A3) in the limit for large energy levels, the normalization factor

$$Z(\beta) = \sum_{n=1}^{\infty} e^{-\beta n}, \tag{A8}$$

which converges for $\beta > 0$, is known as the partition function, so named because it expresses the partition of energies over all possible states the system particles have at disposal, where particles and states stand here for cities and population classes, respectively. The special role played by the partition function lies in the fact that any other thermodynamic variables can be expressed by means of (A8). For instance, the internal energy (A5) can be extracted from the partition function as follows:

$$E = \langle n \rangle = -\frac{\partial \ln Z(\beta)}{\partial \beta}, \tag{A9}$$

which shows that (A8) acts as the generating function of the average energy of the system. Furthermore, substitution of distribution (A7) in entropy (8) gives the maximum entropy value S as follows:

$$\langle n \rangle - TS = -T \ln Z(\beta), \tag{A10}$$

where we exploited relation $T = 1/\beta$. Since $E = \langle n \rangle$, the LHS of (A10) provides the Helmholtz free energy, which represents the work the system can operate onto the environment. Up to the Boltzmann constant, it follows from (A10) that the statistical-mechanical expression for this function is given by:

$$F = -T \ln Z(\beta). \tag{A11}$$

It is worth highlighting that equations (A7)–(A11) come from assuming the first-order moment of the population n as the driving force for a precomplex system of cities to move toward an equilibrium state. This is a straightforward consequence of the Boltzmann hypothesis, according to which the system is assumed to be immersed in an environment that only plays the role of a thermal reservoir, which ensures the temperature to keep fixed, while energy may fluctuate, but the details of such an environment are totally irrelevant to the system equilibrium. On the other hand, the energy constraint (A5) has been introduced in the article from an evolution perspective too. The mean city size indeed gives the characteristic population scale which controls the city system organization when the probability Π_i that a person choses to live in a city i is assumed to be uniformly distributed. This hypothesis is thus coherent with an unstructured thermal environment whose components interact as much randomly as possible. Thus, from a thermodynamic perspective, the energy amount of a pre-complex system of cities is peaked at the characteristic population scale $\langle n \rangle$, while energy levels higher than the mean one exponentially decay.

Entropy competition, and thermodynamic analogy

In this work, far from acting like a simple thermostat, we assumed the environment to be itself a well-structured statistical ensemble, precisely represented by the citizen ensemble G_N, whose structure remains tied to that of the actual system G_A through the constitutive relationship (6). One thus expects the equilibrium configuration to be reached by the system, which we named complex in this case, be totally different from that described by the Boltzmann exponential distribution. In particular, as discussed in the article, when a system of citizens acts as an environment for an ensemble of cities, the phenomenon of entropy competition that emerges in these circumstances allows us to set the logarithmic expectation value of the population

$$\sum_n p_n \ln n = \langle \ln n \rangle_p, \tag{A12}$$

as the least biased thermodynamic constraint to use for maximizing entropy (8), in addition to the normalization condition (A3) (see equations 12). In fact, since the information function for any probability is defined as minus the logarithm of the probability itself, Equation (A12) formally represents the cross-entropy function that gives the average value over the city probability space of the information related to the prior probability $p_n = 1/n\Lambda_2$, which in turn corresponds to the uniform citizen distribution $r_n = 1/\Omega$, according to the unconstrained maximum value of the entropy (9). The Lagrange functional then assumes the following expression:

$$\mathcal{L}\left(p_n\right) = -\sum_n p_n \ln p_n - \lambda \sum_n p_n + \lambda - \gamma \sum_n p_n \ln n + \gamma \langle \ln n \rangle_p, \tag{A13}$$

where γ is the Lagrange multiplier associated to the constraint (A12). Equating to zero the first derivative of (A13) with respect to p_n leads to the following power-law city size distribution:

$$p_n = \frac{n^{-\gamma}}{\zeta(\gamma)}, \tag{A14}$$

where the normalization factor

$$Z_p = \zeta(\gamma) = \sum_{n=1}^{\infty} n^{-\gamma} = \sum_{n=1}^{\infty} e^{-\gamma \ln n}, \tag{A15}$$

derived by using condition (A3) in the limit for large clusters again, is the Riemann zeta function that converges for $\gamma > 1$. In (A15), in order to highlight that $\ln n$ acts as a thermal energy per population class and the power law exponent stands for a generalized inverse temperature $T_p = 1/\gamma$, the Boltzmann factor $e^{-\gamma \ln n}$ has been made explicit. Thus, while the sum over n again must be understood to be a sum over all energy levels the cities can assume, the energy amount per level is no longer given by the population n but its logarithm $\ln n$. It follows the logarithmic mean of the population plays the role of internal energy for a complex system of cities:

$$E_p = \langle \ln n \rangle_p = -\frac{\partial \ln Z_p}{\partial \gamma}, \tag{A16}$$

where the LHS just returns to the constraint (A12), while the RHS shows how the partition function acts as the generating function of the internal energy once more. Note that in the limit for $\gamma \to 0$, (A14) reduces to the uniform distribution $p_n = 1/\Omega$ as well. Moreover, in a manner similar to what was done for the exponential case, knowing the partition function enables us to achieve the following expression for the free energy of a complex system of cities:

$$F_p = -T_p \ln Z_p, \tag{A17}$$

which is understood to hold up to the Boltzmann constant, again.
Finally, by making use of the relation (6), the citizen distribution by population classes corresponding to the most probable city size distribution (A14) can be shown to follow the power law

$$r_n = \frac{n^{1-\gamma}}{\zeta(\gamma-1)}, \tag{A18}$$

which as a function of the scale exponent γ is delayed by 1 with respect to the city size distribution. As above, the partition function for a complex system of citizens in a thermal equilibrium at a temperature $T_r = 1/(\gamma - 1)$ is given by the normalization factor

$$Z_r = \zeta(\gamma-1) = \sum_{n=1}^{\infty} n^{1-\gamma} = \sum_{n=1}^{\infty} e^{-(\gamma-1)\ln n}, \tag{A19}$$

where making explicit the Boltzmann factor again enables us to recognize that the internal energy for a complex system of citizens is defined by

$$E_r = \sum_n r_n \ln n = \langle \ln n \rangle_r = -\frac{\partial \ln Z_r}{\partial \gamma}, \tag{A20}$$

where the subscript refers to the fact the logarithmic mean is calculated on the basis of the r_n distribution. As usual, up to the Boltzmann constant, from the partition function (A19) the following expression comes for the free energy:

$$F_r = -T_r \ln Z_r. \tag{A21}$$

References

Arbia, G. (2001). “The Role of Spatial Effects of the Empirical Analysis of Regional Concentration.” *Journal of Geographical Systems* 3(3), 271–81. https://doi.org/10.1007/PL00011480

Arshad, S., S. Hu, and B. N. Ashraf. (2018). “Zipf’s Law and City Size Distribution: A Survey of the Literature and Future Research Agenda.” *Physica A* 492, 75–92. https://doi.org/10.1016/j.physa.2017.10.005

Batty, M. (1974). “Spatial Entropy.” *Geographical Analysis* 6, 1–31.

Batty, M. (1976). “Entropy in Spatial Aggregation.” *Geographical Analysis* 8, 1–21.

Batty, M. (2010). “Space, Scale, and Scaling in Entropy Maximizing.” *Geographical Analysis* 42, 385–421.

Bogacz, L., Z. Burda, and B. Wacław. (2006). “Homogeneous Complex Networks.” *Physica A* 336(1), 587–607. https://doi.org/10.1016/j.physa.2005.10.024

Cottineau, C. (2017). “MetaZipf. A Dynamic Meta-Analysis of City Size Distributions.” *PLoS One* 12(8), e0183919. https://doi.org/10.1371/journal.pone.0183919

Cristelli, M., M. Batty, and L. Pietronero. (2012). “There is More than a Power Law in Zipf.” *Scientific Reports* 2(1), 812. https://doi.org/10.1038/srep00812

Dorogovtsev, S. N., J. F. F. Mendes, and A. N. Samukhin. (2003). “Principles of Statistical Mechanics of Random Networks.” *Nuclear Physics B* 666(3), 396–416. https://doi.org/10.1016/S0550-3213(03)00504-2

Enel X & Here - City Analytics. Mappa di Mobilità, Retrieved July 1, 2020 from. https://enelx-mobilityflowanalysis.here.com/dashboard/ITA/info.html

Giesen, K., and J. Südekum. (2011). "Zipf's Law for Cities in the Regions and the Country." *Journal of Economic Geography* 11(4), 667–86. https://doi.org/10.1093/jeg/lbq019

Gligor, L., and M. Gligor. (2008). "The Fractal City Theory Revisited: New Empirical Evidence from the Distribution of Romanian Cities and Towns." *Nonlinear Dynamics, Psychology, and Life Sciences* 12(1), 15–28.

Hazen, A. (1914). "Storage to Be Provided in Impounding Reservoirs for Municipal Water Supply." *Transactions of the American society of Civil Engineers* 1308(77), 1547–50.

Istituto Nazionale di Statistica (2019). Annuario Statistico Italiano 2019. Roma. Retrieved from https://www.istat.it/it/files/2019/12/Asi-2019.pdf

Jaynes, E. T. (1957). "Information Theory and Statistical Mechanics." *Physics Review* 106, 620. https://doi.org/10.1103/PhysRev.106.620

Moura, N. J., and M. B. Ribeiro. (2006). "Zipf Law for Brazilian Cities." *Physica A* 367, 441–8. https://doi.org/10.1016/j.physa.2005.11.038

Newman, M. E. J. (2005). "Power Laws." *Pareto distributions and Zipf's law* 46, 323–51. https://doi.org/10.1080/00107510500052444

Rastvorstseva, S., and I. Manaeva. (2016). "Zipf's Law Appearance in the Russian Cities." *Reg. Sci. Inquiry* 8(1), 51–9.

Sanchirico, A., and M. Fiorentino. (2008). "Scale-Free Network as Entropy Competition." *Physical Review E* 78, 046114. https://doi.org/10.1103/PhysRevE.78.046114

Shannon, C. E. (1948). "A Mathematical Theory of Communication." *Bell System Technical Journal* 27(3), 379–423. -10.1002/j.1538-7305.1948.tb01338.x

Singh, V. P. (2013). *Entropy Theory and Its Application in Environment and Water Engineering*. Singapore: Wiley.

Wilson, A. G. (1967). "A Statistical Theory of Spatial Distribution Models." *Transportation Research* 1, 253–69.

Zipf, G. K. (1949). Human Behavior and the Principle of Least Effort. Reading: Addison-Wesley.