International Conference "Water, Environment, Energy and Society" (WEES–2009)
New Delhi, 12–16 January 2009

Mutual Control between Nodes and Edges of a Complex Network

Mauro Fiorentino and Antonio Sanchirico
Dipartimento di Ingegneria e Fisica dell'Ambiente, University of Basilicata
Viale dell'Ateneo 10, 85100 Potenza, ITALY

ABSTRACT: One of the most fascinating discoveries of the last years is that numerous complex systems, although very different among them, as belonging to physical, biological, chemical, informational, social or other environments, exhibit common properties as the scale invariance (i.e. the system properties appear identical over a large range of scales). Moreover, these properties significantly control their dynamics: for instance, it is well known that the scale-invariance of complex networks implies their robustness under random deletion of elements. In such systems, the basic characteristics are not described by the nature of their constitutive elements, but rather by their topological properties (i.e. relations among the system elements). This theory schematizes an interconnected system by a graph, defined as a mathematical set of N nodes (elements of the system) connected by links or edges (relations among the elements). Indeed the node-links schematization allows an effective descriptions for an extremely varied class of phenomena: social networks, as scientific collaboration networks, informatics systems, as the WEB and internet, biological systems, as protein-protein interactions networks and metabolic networks, technological systems, as electronic circuits, and so on. Here, we discuss how such systems self-organize themselves into a steady scale-free structure. In particular, we show that the power-law is the most probable distribution that both nodes and links, in a reciprocal competition, assume when the respective entropy functions reach their maxima under mutual constraint. The proposed approach predicts scaling exponent values in agreement with those most frequently observed in nature.

INTRODUCTION

Network structures are everywhere in nature [1, 2], representing the unifying theme for understanding a number of very different phenomena across social, biological, technological, hydrological and ecological systems. A complex network, also called graph in the mathematical literature, may be simply defined by N nodes (or vertices) connected by M edges (or connections); nodes are system elements, while edges represent their interactions.

The usefulness of the network theory lies in its universality, indeed the node-edges schematization allows for an effective description for an extremely varied class of phenomena. For example, Internet is a network of computers and routers connected by a physical and wireless link [3, 4]; likewise the World-Wide-Web is a virtual net of HTLM documents, connected by hyperlinks pointing from one document to another [5]. The cell metabolism is guaranteed by a complex network of substrates and enzymes, connected by chemical reactions in which these molecular compounds participate [6]. The food webs may be thought of as directed networks, where nodes are distinct species and edges represent predator-prey relationships [7]. News and diseases spread on a social network, where vertices are human beings and edges represent various social interactions among them [1, 8, 9]. As regards technological systems, the power grid can be described by a network of generators and transformers connected by high-voltage transmission lines [1, 9].

In all these cases the representation of the system as a network is obvious, nodes and links being identified directly. In other cases, instead, the mechanisms that couple the interacting nodes may be more complex and the edges can be identified by noting similarities in the dynamical behaviour of two nodes [10]. Recently, based on this idea and analyzing daily temperature records taken from a grid in various geographical zones, Yamasaki *et al.* [11] developed a method for generating climate networks, where edges represent correlations between temperature in different sites. In particular, the authors find that the dynamics of the network of temperature are significantly influenced by El-Niño, even in geographical zones in the world where the mean temperature level is not affected. Thus, fast fluctuations of the correlations observed during El-Niño periods cause the links to break, generating strong climate changes. Moreover, a soil pore network can be useful in studying soil moisture patterns [12]. In this network centres of pores represent nodes, whose connectivity may be quantified through the number of independent paths between two points. Likewise, complex networks of fractures in rocks have been

studied in [13]. Furthermore, with reference to the geomorphology, river networks may be defined by nodes on a regular lattice, representing the elevation field, while links are determined by steepest descent on the topography [14].

The most schematic characteristic of a graph is the node degree k_j (also called connectivity), defined as the number of connections of node V_j. A node V_j is said to be a nearest neighbour of V_i if there is an edge between these two nodes. Therefore, the degree of a node is actually the number of its nearest neighbours. Fixed N, the degree sequence $D_N = \{k_1, k_2, \ldots, k_j, \ldots, k_N\}$ is a set of integers, whose elements stand for the connectivities of all nodes of the graph [15]. The degree distribution, that is the probability that a randomly selected node has degree k, is given by $p_k = N_k/N$, where N_k is the number of nodes with k edges.

A multitude of studies have shown that most real-world complex networks display a strongly heterogeneous topology, described by a power-law degree distribution $p_k \sim k^{-\gamma}$, which implies that the highly connected nodes are statistically significant [1, 2]. In all networks considered the values of the scale exponent γ typically range from 2 to 3 [2], although extreme values of about 1 (food webs [7]) and 3.54 (net of human sexual contacts [16]) have been observed. The shape of the power-law degree distribution implies a continuous hierarchy of nodes, going from a few vertices with a high degree to a number of nodes with very few edges. Thus, the degree distribution displays a long-tail behaviour that begets the extreme values (hubs). Due to the lack of a characteristic scale, these networks are also called scale-free networks [1, 2], satisfying the property $p_{ak} = g_a p_k$. Indeed, as reviewed by Newman [17], if we increase the scale by which one measures k by a factor α, the shape of the distribution does not change, except for the overall multiplicative constant g_α. Thus, the power-law distribution can be taken as a synonym for the notion of scale-free or scaling law, in the sense that the system properties appear identical under a scale change. Furthermore, the most actual usage of the term scale-free appears to have a richer sense, being associated with additional features, such as self-similarity and fractal geometry [18, 19].

Scaling laws and fractals are abundant in nature and river networks stand for a fascinating epitome of such phenomenon [20]. Indeed, despite great diversities in the geologic, vegetational and climatic constraints, river networks self-organize into recurrent patterns, showing a number of topological regularities associated with their tree-branching structure [20–23]. Furthermore, fluvial networks display the absence of a single well defined length scale too, reflecting in a power-law distribution of various quantities [20–22], such as the distributions of stream lengths and basin areas [22]. Consequently, a comprehensive understanding of the spatial (and temporal) patterns describing the networked system may be fundamental to many areas of earth sciences, such as geology, geomorphology and hydrology, providing much greater insights into dynamical processes than traditional methods [24]. For instance, in Hydrology it is well known [25] that network structure and specific system functions co-evolve with time, with strong interactions and feedbacks between patterns and processes at both catchment and fluvial scale. Furthermore, in other fields, it has been proved that the topology of the complex networks affects their behaviour in a fundamental way. Particularly, it has been demonstrated that the networks with a power-law degree distribution display a surprising robustness against random removal of their own elements [26]. In biology this property explains the strong resilience exhibited by simple organisms under random deletion of genes [27].

Complex networks can be substantially studied by adopting two complementary approaches [28]. One on side, some models [1, 29–31] deal with non-equilibrium growth networks, simulating the time evolution of a graph and reproducing the dynamics responsible for its topology. An example of such models is that proposed by Barabási and Albert [1], based on preferential attachment rule (while a network grows, new nodes attach preferentially to the already well connected nodes). This model asymptotically produces a stationary and power-law degree distribution with scaling exponent $\gamma = 3$. Several refined variants of preferential attachment have been proposed, giving a more realistic description of the local events that control the networks evolution. Besides the addition of new nodes, these dynamical models allow for the addition and rewiring of the links [29], the initial attractiveness of the nodes [30], and the nonlinear preferential attachment [31]. Depending on the weights of these processes, the scaling exponent of the power-law continuously changes between 2 and ∞, accounting for the large variation observed in nature. Note that, in other fields, the preferential attachment mechanism resembles some dynamics introduced in order to simulate eco-hydrological processes. In facts, recently, using analyses of satellite image data, two groups of researchers [32] have demonstrated that patch sizes for vegetation ecosystems in arid and semi-arid climatic

zones obey power-law distributions. Both groups proposed a stochastic model for vegetation patterns, wherein each vertex of a lattice is either occupied or unoccupied and the number of old (new) shrub that die (sprout) is proportional to the population of the neighbour vertices. Then, the chance of a large cluster to grow is larger than that of the small clusters and, consequently, the distribution of the cluster sizes becomes wide, thus obeying a power-law.

Complex networks have been investigated via various optimization methods, by assuming that the networks evolve in order to perform some specific functions. For instance, in [33] scale-free networks are obtained by minimizing a linear combination of the link density and average distance (i.e., shortest path between two nodes). Likewise, the results found in [34] suggest that the emergence of the scale-free topology may arise from a tradeoff between the requirement of having the shortest route between any two nodes and the smallest congestion in the network. Moreover, in [35], by minimizing a cost function depending both on length and traffic carried by the edges, the authors proposed a model of networks, whose topology display a spanning tree shape, thus showing a spatial hierarchical organization.

Furthermore, network optimization is actually known to be relevant in the study of general transportation networks. Indeed, Banavar *et al.* [36] showed that the network topology arises from an optimization process, in which edges are positioned to minimize a specific quantity, representing energy or cost of transportation. In particular, the authors showed that trees (i.e., networks without loops) represent configurations of local optimum for the specific functional to be minimized, thus, from an evolutionary viewpoint, trees prevail over networks [36]. This result is very general, suggesting that other natural tree-like structures can arise through optimality to different selective pressures. For instance, this result can explain the allometric scaling in biology [37], because of that many biological processes, from cellular metabolism to population dynamics, are characterized by a power-law relationship between size (body mass) and metabolic rate. Similar scaling behaviour is obtained for river morphology too, wherein the scale-invariant tree-like patterns of the river networks can be though of as the natural by-product of the minimization of the total energy dissipation [38], thus providing a unifying approach for biology and hydrology.

These results suggest that variational approaches represent a valid tool accounting for the observed regularities displayed by most real-world complex networks. On the other hand, systems like the complex networks, consisting of a number of interacting elements, naturally lend themselves to a statistical description, thus a statistical mechanics approach seems to be appropriate. This basic idea has led to a new class of models [28, 39–41] that use standard tools of the equilibrium statistical mechanics. In particular, in order to investigate uncorrelated scale-free networks at equilibrium state, Sanchirico and Fiorentino [41] have introduced the novel concept of entropy competition, by means of which a power-law degree distribution is achieved. In this paper, using tools of the statistical physics and emphasizing the equivalence between entropy and information (especially with regard to the statistical inference), we review how three classes of degree distributions can be derived by means of the maximum entropy principle, according to the role played by the connectivity of the nodes. We start from the review of a statistical-mechanical model of literature, which plays in network theory the role that the Boltzmann model plays in statistical mechanics. Then we develop a consistent formalism based on equilibrium statistical mechanics. According to the Maxwell-Boltzmann statistics, we achieve an exponential degree distribution, by considering the connectivity as an indefinite property of the system elements, without any reference to the networked structure. Moreover, as a test of our approach, we derive the Poissonian degree distribution, by thinking of the connectivity as a topological property, coinciding with the number of half-edges going out from a node. Further we explain how two entropy functions can reach their maxima, under mutual constraints, at which the degree distributions of both nodes and first neighbours follow a power-law. In this way, we point out new insights into topological quantities that control the scaling exponent and there after we conclude.

STATISTICAL MECHANICS OF COMPLEX NETWORKS

Different statistical mechanics approaches have been used to characterize complex (scale-free and finite-scale) networks [28, 40]. The underlying models define statistical ensembles of networks, endowed with a probability measure, constituting a configuration space (the phase space). Then, the topological features of the networks can be derived by an appropriate choice of the probability measure. It is worthwhile to review the theoretical back-ground underlying the Exponential Random Network (ERG) model, introduced by Park and Newman [39].

Exponential Random Graphs and Configuration Model

The ERG model assumes as a statistical ensemble, $\mathcal{G}$, a collection of simple graphs (i.e. graphs without self-connections and multiple edges), and defines their probability distribution by maximizing entropy under appropriate constraints, imposed by a given set of observations. In so doing, the choice of the probability distribution is not affected by any a priori bias, as the only reliable information is enclosed in the constraints for entropy maximization.

An observable y_G is defined as a quantity that describes some property of the elements of $\mathcal{G}$. For instance, an observable may be the number M of edges or the degree sequence D_N of a graph. The estimate $\langle y_G \rangle$ of the expectation value of a given observable, deduced by measurements performed over real networks, can be used to derive the probability distribution p_G, with $G \in \mathcal{G}$, by imposing that this estimate equals the mean $\mathrm{E}[y_G] = \sum_G p_G y_G$ over the statistical ensemble. In other words, we want to choose a probability distribution p_G such that the networks that better fit the observed characteristics $\langle y_G \rangle$ have statistical weight p_G, thus satisfying constraint $\sum_G p_G y_G = \langle y_G \rangle$. As a rule for the definition of statistical weight, the ERG model assumes that the best choice of the probability distribution p_G is the one that maximizes entropy, defined as $S_G = -\sum_G p_G \ln p_G$, under the normalization condition $\sum_G p_G = 1$, and constraint $\sum_G p_G y_G = \langle y_G \rangle$. Such a maximization reproduces the conditions of maximum randomness possible, according to the chosen constraints, which represent the only reliable information on the system state. Using the Lagrange multipliers method, Park and Newman [39] demonstrated that the solution of the above constrained maximum problem is given by the Boltzmann exponential distribution $p_G = \mathrm{e}^{-H(G)}/Z$, were the function $H(G) = \mu y_G$ (with μ Lagrange multiplier) is called Hamiltonian of the network. The normalization constant $Z = \sum_G \mathrm{e}^{-H(G)}$ is called partition function, borrowing the term from statistical mechanics. Constant μ can be interpreted either as a field coupling to the observable or as an inverse temperature [39]. Particularly, $\langle y_G \rangle = -\partial F/\partial \mu$, where the function $F = -\ln Z$, called free energy, stands for the generating function of the expectation value of the observables. So, the observable y_G plays in the theory of the exponential random graphs a role similar to the energy in statistical mechanics. Using this formalism, and assuming the number M_G of edges in the graph as only observable y_G, Park and Newman [39] obtain the Poissonian degree distribution, $p_k = \mathrm{e}^{-\mathrm{E}[k]}\mathrm{E}[k]^k/k!$, previously derived by Erdös and Rényi [42], by randomly connecting with equal and independent probability p each of the $N(N-1)/2$ possible pairs of nodes in a network of N vertices ($\mathrm{E}[k] = pN$). Note that the Poissonian degree distribution has a maximum at the average degree $\mathrm{E}[k] = \sum_k k p_k$ of the network, implying that each node has approximately the same number of links, whereas the probability of finding a node with connectivity greater than the mean one decays exponentially.

On the other hand, the ERG model can be generalized assuming as an observable a given degree sequence $D_N = \{k_1(G), \ldots, k_N(G)\}$ for each graph of $\mathcal{G}$ [39]. Thus, one obtains graphs that, in all respects other than their degree sequence, are assumed to be entirely random. The idea to simulate the randomness of the Poissonian graphs while satisfying a given degree sequence D_N has led to a new type of model, known as configuration model (CM) [15, 43, 44], that extends the Erdös-Rényi model to the graphs with degree distribution that differs from the Poissonian one.

The CM can be defined by means of the following heuristic procedure described by Newman [44]: First we assign to each node V_j a number k_j of ends of connections (half-edges) drawn from the set D_N, under the constraint that the sum $\sum_j k_j$ is even ($1 \leq j \leq N$). If $\sum_j k_j$ is odd, then one can simply add a vertex of degree 1 without changing the topological properties of the graph. Next, we randomly choose pairs of half-edges and join them together to make complete edges, thus building up a real graph. Nevertheless, a drawback of this model is that it produces multigraphs, as the construction rules do not avoid multiple edges and self-loops. Therefore, in order to implement the CM, it is critically important to determine the conditions by means of which a random configuration defines a simple graph. To this end, since we will analyze the behaviour of graphs in the limit of a large system size, some technical conditions of uniform convergence must hold (see [15] for more details). In particular, as stated by Molloy and Reed [15], the main result in [45] implies that the underlying multigraph of a random configuration with a given degree sequence and maximum degree $K = O(N^{1/4})$ is simple with probability tending to a constant when both $E[k] = \sum_k k p_k$ and $Q(D) = \sum_k k(k-2)p_k$ are finite. In the case of the power-law degree distribution $Q(D)$ tends to infinity for $2 < \gamma \leq 3$, whereas in the range $1 < \gamma \leq 2$ both $Q(D)$ and $E[k]$ diverge. Nevertheless, these quantities may assume finite values for all γ, as showed by Aiello *et al.* (see [46] for more details), using a cut-off like $K = O(N^{1/\gamma})$. The cut-off condition, is needed to reduce the multiple

edges, as the probability to have multiple connections increases when the system size grows. Among other things, the degree distributions of many real-world networks are truncated on the right [9], thus displaying a cut-off at higher degree values, due to the finite system size [47]. Several authors use different conditions of cut-off than that used by Aiello *et al.* For example Newman *et al.* [44] assume an exponential cut-off, whereas Dorogovtsev *et al.* [48] define the natural cut-off of the network as $K = O(N^{1/\gamma-1})$. Although the cut-off required in [15] is much smaller than the natural one, for scale-free networks with scaling exponent between $2 < \gamma \leq 3$ (i.e., for γ in the range of values more frequently observed in nature), in the limit of large system size, the fraction of multiple edges over the total number of links vanishes, using the structural cut-off $K = O(N^{1/2})$ introduced by Boguñá *et al.* in [47]. Consequently, any result that holds for a power-law configurations with $2 < \gamma \leq 3$ and $0 \leq k \leq N^{1/2}$ does for the underlying simple graphs too.

We underline that the configuration model provides only a particular realization of all graphs on N vertices and with degree sequence D_N. Each of these graphs is a random member of the set $\mathcal{G}(D_N)$ of all graphs on N vertices and with fixed degree distribution p_k. Hence, we may think of $\mathcal{G}(D_N)$ as a microcanonical ensemble. Moreover, since the graph is constructed by joining pairs of half-edges randomly selected, the degrees of all vertices are statistically independent. So, the configuration model is not capable of simulating the significant correlations among the degrees of the neighbour nodes of real-world networks, [4, 49].

MAXIMUM ENTROPY PRINCIPLE

Although statistical mechanics has been developed to describe the properties of the gases, its field of application has been extended to the study of fluid, solid state and other physical systems. In this Section, reviewing the results in [41], we explain a more direct analogy than the one used by ERG models. In the following, we will think of the elements (nodes) of a complex network as the elementary particles (atoms or molecules) of a gas. So, the extension of statistical physics laws to the complex networks becomes more natural. In particular, we assume as a statistical ensemble (denoted by G_n) the collection of all nodes of a graph, and we associate to each node a certain statistical weight. In this way, the maximum entropy principle can be employed to achieve the most probable degree distribution. We underline that the ensemble G_n is simply a collection of nodes but not a proper graph. Nevertheless, once a given degree sequence, marking our G_n ensemble, has been determined, the shift from G_n to the real graph G becomes immediate by using the configuration model described in the previous section, once required hypotheses have been verified.

Exponential Degree Distribution

Given a graph, each node V_j is marked by the value k_j of its connectivity. So, all nodes that have the same degree can be grouped together in a certain number of clusters (say Ω). Specifically, assuming for simplicity that the degree sequence D_N with the largest element K is full (i.e., each integer k satisfying $0 \leq k \leq K$ is an element of D_N), the number of clusters Ω is equal to $K + 1$. Moreover, in the limit of large system size, a cut-off like $K = \Theta(N^{1/2})$ ensures that the number of clusters is always less than that of nodes. It is well known that the number W_1 of ways to arrange the N nodes (suppose identical, distinguishable and independent) of G_n in Ω clusters, with the kth cluster containing N_k nodes of connectivity k, is given by the multinomial coefficient,

$$W_1 = \frac{N!}{\prod_{k=0}^{K} N_k!} \quad \ldots (1)$$

In this case, the kth cluster, identified by the value of the connectivity of its nodes, represents the kth cell of the phase space of the statistical ensemble G_n. For large networks, in analogy with statistical mechanics, the entropy function of this system (that we denote by S_p) can be defined as proportional to the logarithm of W_1 [41]. Thus, omitting a proportionality constant, we can write,

$$S_p = -\sum_{k=0}^{K} p_k \ln p_k \quad \ldots (2)$$

where, the frequency $p_k = N_k/N$ stands for the probability that a randomly selected node lies in the kth cell of the phase space. From a statistical mechanics point of view, S_p measures the disorder encoded in the degree distribution, quantifying the heterogeneity of the node degree.

On the other hand, entropy (2) also admits an interpretation in terms of the information theory. To this end, let us consider an experiment $X = \{x_1, x_2, \ldots, x_k, \ldots, x_\Omega\}$, with Ω possible outcomes, defined by the occurrence probability $p_k = N_k/N$, where N is the total number of results and $N_k = |\{X: X = x_k\}|$ denotes the number of times in which the experiment X has given the result x_k. In the context of information theory, random variable X may be thought of as a message transmitted by a discrete source of information, and p_k

represents the probability that this message is implemented by symbol x_k. Shannon [50] proposed as a measure of the uncertainty of a message transmitted by a source the following quantity $H(X) = -\sum_k p_k \ln p_k = \mathrm{E}[-\ln p_k]$, with $1 \le k \le \Omega$, called entropy of message and coinciding with the Boltzmann-Gibbs entropy (2). The quantity $I(p_k) = -\ln p_k$ is called information related to the probability distribution p_k. In a statistical context, like that of the information theory, $H(X)$ measures the uncertainty that the experimenter has before performing the experiment. If the only information on random variable X consists of an estimate $\langle f(x) \rangle$ of the expectation value of one function $f(x)$ (or more), then we have a problem of inference from incomplete information. As suggested by Jaynes, the maximization of the entropy (2) simulates the conditions of maximum uncertainty, due to the partial knowledge of the processes that generate the random variable [51]. This kind of approach does not contaminate the statistical inference with preconceived hypotheses on the features of the random variable. On the contrary, this method allows us to build a probability distribution over a given statistical ensemble on the basis of partial knowledge of the system, the only information being given by the constraints for the entropy maximization. So, the maximum entropy method provides the best estimate (in the sense of least biased) with regard to missing information. Thus, as a thermodynamical system achieves the maximum disorder at equilibrium state, the most probable distribution of a statistical ensemble is the one that maximizes uncertainty. Then, with regard to complex networks, entropy S_p, coinciding with that defined by Shannon in the information theory, may be thought of as a measure of the uncertainty related to the topological configuration of a graph. This entropy measures how random is the degree distribution of a complex network.

It is well known [52] that by maximizing the entropy S_p only under the normalization constraint $\sum p_k = 1$, one obtains the following uniform distribution,

$$p_k = \frac{1}{\Omega} \qquad \ldots (3)$$

which implies that each cluster of G_n is equally likely. Indeed, in Eqn. (3) the probability p_k that a randomly selected node lies in the kth cluster also represents the probability to randomly select a given cluster of G_n. Let us denote with p_j the probability associated with the generic node V_j of the statistical ensemble, and randomly choose a node of G_n. This choice can be performed first by choosing the cluster containing the node and next selecting one of the N_k nodes of such cluster. Holding Eqn. (3), we can write $p_j = 1/N_k\Omega$, where with $1/N_k$ we have denoted the probability to randomly select one of the N_k equivalent nodes of the kth cluster. Since for Eqn. (3) all clusters are equally likely (i.e., they have the same number of nodes), one has $N = N_k\Omega$ and so the probability p_j becomes,

$$p_j = \frac{1}{N} \qquad \ldots (4)$$

Thus, each node of G_n has the same probability p_j to be chosen, independently of its own connectivity. Since both equations (3) and (4) assign identical weights to all elements of the system, they equally provide a microcanonical description of the statistical ensemble G_n (supposed isolated).

Once the microcanonical ensemble has been established, the next step in a statistical-mechanical formulation is to choose the constraints to impose on the system, thus defining the partition function, the temperature and the canonical ensemble. Since each cluster is identified by a different value of the connectivity, it is quite natural to contemplate the mean degree, $\mathrm{E}[k] = \sum_k k p_k$, of the network for entropy maximization. So, assuming the degree k as an observable, we maximize entropy S_p under the normalization condition $\sum_k p_k = 1$, plus the other constraint $\sum_k k p_k = \langle k \rangle$, thus assuming that the mean degree $\mathrm{E}[k]$ equals the estimate $\langle k \rangle$ of the expectation value of the connectivities. Using the Lagrange multipliers formalism, one obtains the exponential degree distribution $p_k = \mathrm{e}^{-\mu k}/Z$ [41], which is to the graphs theory as Boltzmann distribution is to the equilibrium statistical mechanics. The normalization constant, $Z = \sum_k \mathrm{e}^{-\mu k} = \mathrm{e}^{\mu}/(\mathrm{e}^{\mu}-1)$, represents the partition function, while the Lagrange multiplier $\mu > 0$ may be though of as an inverse temperature [39]. In analogy with statistical mechanics, with $H(k) = \mu k$ and $F = -\ln Z$ we denote the Hamiltonian and the free energy, respectively. Since constraint $\sum_k k p_k = \langle k \rangle$ may be written as $\langle k \rangle = \partial \ln F/\partial \mu$, by calculating the first derivative of F one has $\langle k \rangle = 1/(e^{\mu}-1)$ that, solving for μ, gives $\mu = \ln(1+1/\langle k \rangle)$, from which it follows that the values of constant μ are always positive for all $\langle k \rangle > 0$.

Following Molloy and Reed [15], in order to find the position of the phase transition for an exponential configuration, we substitute the exponential degree distribution $p_k = \mathrm{e}^{-\mu k}/Z$ in the inequality $\sum_k k(k-2)p_k > 0$, thus obtaining $(\partial^2 Z/\partial \mu^2 + 2\partial Z/\partial \mu)/Z > 0$, that is satisfied for values of $\mu < \ln 3$. Consequently, in the case of exponential degree distribution, the phase transition of the underlying graph occurs when $\mu = \ln 3$ or, equivalently, when $\langle k \rangle = 1/2$. Moreover, both $\mathrm{E}[k]$

and $Q(D)$ are always finite for all $\mu \neq 0$, and so the underlying multigraph has a non-zero probability to be simple.

Although the tail of the exponential distribution rapidly decays, as it happens for the Poissonian graphs, an exponential degree distribution well simulates the behaviour of some real-world networks, among which the neural network of the nematode *C. elegans*, whose neurons stand for the nodes and the axons are the edges of the network [9]. Also these systems have been explained by dynamical modes, by connecting at random new nodes added to the graph (i.e., maintaining the growth hypothesis but removing the preferential attachment rule) [1]. On the contrary, here we have shown how an exponential degree distribution may be derived from a general principle, linking the sole parameter, μ, of the distribution to the average degree of the network.

Finally, since in the limit for $\mu \to 0$ (high temperatures) $Z \to \Omega$, the Boltzmann degree distribution reduces to Eqn. (3), and, consequently, the canonical ensemble would coincide with the microcanonical one. To complete the chapter devoted to the general considerations, in the next section we will test our approach in the case of classical random graphs.

Poissonian Degree Distribution

As discussed in the previous subsection, each node V_j has a certain value k_j of the connectivity, which has been thought of as a discriminating characteristic of the nodes, without any topological meaning. In other words, the statistical description interprets k_j as a property (called connectivity) of a node, and by means of this property we can distinguish one node from another, identify the ones of the same type, and group them in Ω clusters. In fact, the connectivity property of a given node V_j is due to a number of edges (equal to k_j) among V_j and its first neighbours. In this subsection, in order to express this topological nature of the connectivities, we consider not the node but the hybrid element, consisting of the node and of its half-edges, as a component of a statistical ensemble, whose construction will be described shortly (see Figure 1).

Since the number of half-edges going out from each node coincides with the value of the node degree, the hybrid elements can be grouped in the same clusters, previously defined on the basis of the different values of connectivities. In this case, in order to build such groups, first we need to arrange in the Ω clusters both N nodes and $2M$ half-edges of the graph, and next to assemble the hybrid elements by assigning half-edges to each node of the given cluster. Note that, once the total number N of nodes has been fixed and for each degree sequence D_N, we do not need to independently specify the number of half-edges $2M$, because (for undirected graphs) one has that $2M = \sum_j k_j = \sum_k k N_k$, for $1 \leq j \leq N$ and $0 \leq k \leq K$.

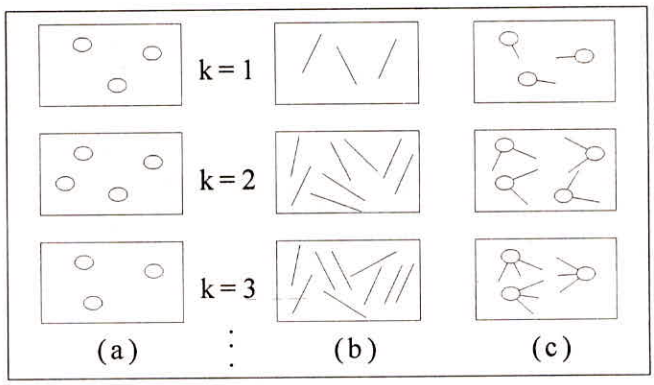

Fig. 1: The construction procedure for the hybrid elements, which consist of a node with its own half-edges. (a) For all k, each cluster contains N_k nodes. (b) Each cluster contains $C_k = N_k k$ half-edges, that is k half-edges for each node of the respective cluster in (a). (c) For each k, we assign the C_k half-edges of a given cluster in (b) to the N_k nodes of the respective cluster in (a)

The construction of the statistical ensemble can be subdivided into three following steps [41]:

Step 1. Assign N_k nodes to each of Ω clusters, where the kth cluster contains all nodes of connectivity k, or rather we will assign (at step 3) k half-edges to the nodes belonging to this cluster (see Figure 1a). In a manner similar to that in Section III A, the number W_1 of ways to arrange N nodes in Ω clusters, with the kth cluster containing N_k nodes, is given by the multinomial coefficient, defined by Eqn. (1).

Step 2. In these same clusters we need to arrange the half-edges too. Since the kth cluster contains N_k nodes (of degree k), we arrange $C_k = N_k k$ half-edges in this cluster, that is k half-edges for each of N_k nodes. For the sake of simplicity of representation, as shown in Figure 1b, we arrange the half-edges in other Ω clusters adjacent to the first ones. Let us denote with $C = \sum_k C_k$ the total number of half-edges (connectivities) of the graph. The number of ways W_2 to arrange C half-edges (supposed identical, distinguishable and independent) in Ω clusters, with the kth cluster containing C_k half-edges, is given by the multinomial coefficient,

$$W_2 = \frac{C!}{\prod_{k=0}^{K} C_k!} \qquad \ldots (5)$$

Step 3. Now, fixed the kth pair of clusters, we need to assign the C_k half-edges to the N_k nodes, thus obtaining the hybrid elements belonging to the kth cluster of Figure 1c. This problem is equivalent to the one of disposition of the C_k objects in the N_k cells to k to k, whose solution is again given by the multinomial coefficient $W_{3,k} = C_k!/k!^{N_k}$. Since, step 3 must be repeated for each pair of clusters, the total number W_3 of ways to assign all half-edges to the N nodes is given by $W_3 = \prod_k W_{3,k}$.

Finally, by using the multiplication principle of combinatorics, the total number W of ways to arrange the N hybrid elements in the Ω clusters is given by,

$$W = \frac{N!}{\prod_{k=0}^{K} N_k!} \frac{C!}{\prod_{k=0}^{K} C_k!} \prod_{k=0}^{K} \frac{C_k!}{k!^{N_k}} = \frac{N!C!}{\prod_{k=0}^{K} N_k! k!^{N_k}} \quad \ldots (6)$$

Again in analogy with the statistical mechanics, we can assume the entropy S of the statistical system as proportional to the logarithm of W. Then, using Stirling's approximation ($N! = N^N e^{-N}$ and $C! = C^C e^{-C}$, for $N \to \infty$ and, consequently, for $C = \mathrm{E}[k]N \to \infty$), after simple algebra, we obtain,

$$S \propto -\sum_{k=0}^{K} p_k \ln p_k k! + \mathrm{E}[k]\left(\ln N + \ln \mathrm{E}[k] - 1\right) \quad \ldots (7)$$

where the frequency $p_k = N_k/N$ stands for the probability that a randomly selected node has k half-edges or rather it is the probability that one of the N hybrid elements of our statistical ensemble belongs to the kth cluster. Note that, if the average degree $\mathrm{E}[k]$ is assumed constant (as we will do in the following), the monotonicity and concavity properties of S are only controlled by the first term on the right hand side of Eqn. (7), as the second term is a constant. Thus, since the summation in Eqn. (7) is a concave functional of p_k, the probability distribution at which the first derivative of S is zero also maximizes S.

Let us maximize entropy (7) under the normalization condition $\sum_k p_k = 1$, plus the constraint $\sum_k k p_k = \langle k \rangle$. The Lagrangian $L(p_k)$ is given by,

$$L(p_k) = -\sum_{k=0}^{K} p_k \ln p_k - \sum_{k=0}^{K} p_k \ln k! + \langle k \rangle \left(\ln N + \ln \langle k \rangle - 1\right) - \lambda \sum_{k=0}^{K} p_k + \lambda - \mu \sum_{k=0}^{K} p_k k + \mu \langle k \rangle \quad \ldots (8)$$

where λ e μ are two Lagrange multipliers. Using constraint $\sum_k k p_k = \langle k \rangle$, we have substituted the average degree $\mathrm{E}[k]$, that appears on the right hand side of Eqn. (7), with its estimate $\langle k \rangle$, by assuming that the sum $\sum_k k p_k$ keeps constant (and equal to $\langle k \rangle$) while p_k changes. Thus, equating to zero the first derivatives of the Lagrangian, we obtain,

$$p_k = A \frac{b^k}{k!} = e^{-b} \frac{b^k}{k!} \quad \ldots (9)$$

where we have set $A = \mathrm{e}^{-(1+\lambda)}$ and $b = \mathrm{e}^{-\mu}$. The latter equality in Eqn. (9) follows from the normalization constraint $\sum_k p_k = 1$, by using the Taylor series expansion of the exponential function around zero. Finally, substituting Eqn. (9) in the constraint $\sum_k k p_k = \langle k \rangle$, one has,

$$\langle k \rangle = \sum_{k=0}^{\infty} e^{-b} \frac{b^k}{k!} k = e^{-b} b \sum_{k=0}^{\infty} \frac{d}{db}\left(\frac{b^k}{k!}\right) = b \quad \ldots (10)$$

thus obtaining [41] the Poissonian degree distribution $p_k = \mathrm{e}^{-\langle k \rangle} \langle k \rangle^k / k!$, introduced in Section II A.

Note that, from the above positions it follows that the Lagrange multiplier μ is given by $\langle k \rangle = \mathrm{e}^{-\mu}$, being a function only of the mean degree, which is finite for all $\mu \neq -\infty$. Finally, the partition function is given by $Z = \mathrm{e}^{\langle k \rangle} = \exp[\exp(-\mu)]$, thus satisfying the condition $\langle k \rangle = \partial \ln F / \partial \mu$.

The Poissonian degree distribution has also been derived by Park and Newman by maximizing the S_G entropy, as well as in other ways by Erdös and Rényi. This is not surprising as all three methods are founded on the hypothesis of maximum randomness. Yet, there are substantial differences between two previous models and our approach: Erdös-Rényi random graph model is not based on statistical mechanics formalism, and in our case we have chosen a statistical ensemble different from the one assumed by ERG model. Particularly, this model assumes as a statistical ensemble a collection of graphs, whereas our method refers to the set of the N hybrid elements previously defined. This result proves the high predictive power of our approach, by showing that statistical ensembles based on degree sequence may be helpful for an in-depth analysis of the networks structure. In the next sections, we will show how a power-law can be thought of as the most probable degree distribution, by means of the concept of entropy competition, [41].

SCALE-FREE NETWORKS

We have obtained a Poissonian degree distribution, thinking of the connectivity as the number of half-edges going out from a node. Really, the nodes of a network are real physical elements, whereas the edges stand for a mathematical schematization that describes

the interactions among the elements of a complex network. In other words, although the edge is a topological element of a mathematical object (graph), with regard to the real-world networks it seems more appropriate to interpret the connectivity of a node as the number of its first neighbours, thus giving a quite natural physical interpretation of the connectivities, which, in this way, are thought of as elements.

Two Statistical Ensembles

Given a graph, each node V_j is linked to other k_j nodes, the so called nearest neighbours of V_j. Consequently, in addition to the statistical ensemble G_n, consisting of N nodes we can define a second statistical ensemble (that we denote with G_{nn}) consisting of all nearest neighbours of the N nodes of the graph. Particularly, once the total number N of nodes has been fixed and for each given degree sequence $D_N = \{k_1, k_2, \ldots, k_j, \ldots, k_N\}$, the sum of the elements of this sequence defines the total number $C = \sum_j k_j = \sum_k N_k k$ of the first neighbours of all nodes (for $1 \leq j \leq N$ and $0 \leq k \leq K$). Thus, each node V_j of the graph, thought of as an element of connectivity k_j, belongs to the statistical ensemble G_n. Instead, if we think of V_j as the first neighbour of other k_j nodes, then V_j will be k_j times an element of the statistical ensemble G_{nn}, one time for each of its neighbour nodes. This last assertion implies that, strictly speaking, G_{nn} is a multiset, containing the same element (V_j) more than once too. But this does not affect the combinatorial counts as all elements of G_{nn} will be considered identical. In Figure 2, we present a graph with 4 nodes together with its two statistical ensembles. Nodes belonging to the clusters of G_n (Figure 2b) have connectivity k, whereas the elements that belong to the corresponding clusters of G_{nn} (Figure 2c) are the first neighbours of the nodes of degree k. Due to this scheme, we can group the N elements of G_n and the $C = \sum_k N_k k$ elements of G_{nn} in Ω pairs of clusters, with the former and the latter clusters of the kth pair containing N_k and $C_k = N_k k$ elements, respectively. Once more, the numbers W_1 and W_2 of ways to perform such arrangements are given by the multinomial coefficients (1) and (5), respectively. Consequently, while entropy S_p of the statistical ensemble G_n is defined by Eqn. (2), the entropy (that we denote with S_r) of the statistical ensemble G_{nn} may be defined as proportional to the logarithm of W_2 [41]. Thus, omitting a proportionality constant, we can write,

$$S_r = -\sum_{k=0}^{K} r_k \ln r_k \qquad \ldots (11)$$

where the frequency $r_k = C_k/C$ is the ratio between the number $C_k = N_k k$ of the nearest neighbours of all nodes of degree k and the total number $C = \sum_k N_k k$ of elements of G_{nn} (i.e., the total number of the first neighbours of all nodes of the graph). In accord with Newman [44], since the probability that a randomly selected edge reaches a node of degree k is proportional to kp_k, for uncorrelated networks, the frequency $r_k = C_k/C = kp_k/\mathrm{E}[k]$ also stands for the degree distribution of the nearest neighbours of a randomly selected node [4, 53]. Thus, the two entropy functions (2) and (11) give a measure of the highly skewed distribution of the connectivities of a graph from a point of view of both nodes and their first neighbours, respectively.

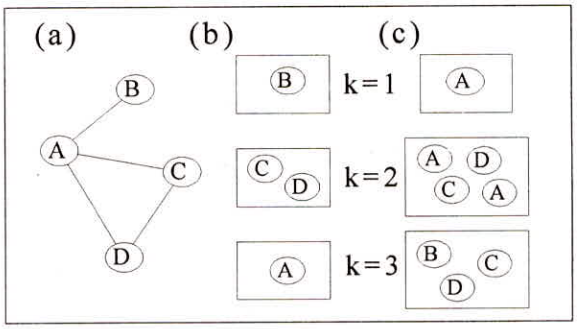

Fig. 2: (a) A graph with $N = 4$ nodes and $C = \sum_j k_j = 8$ nearest neighbours. (b) Three clusters of G_n, containing all nodes of the graph. (c) Three clusters of G_{nn}, whose elements are the first neighbour of the nodes of G_n. For instance, as the node V_C has degree 2, it belongs to the second cluster of G_n. The same node also belongs to both second and third cluster of G_{nn}, being the first neighbour of both nodes V_D and V_A, that have degree 2 and 3, respectively

We note that, since $\ln W_2 = -\sum_k C_k \ln r_k$, entropy S_r should be defined as $S_r = -\mathrm{E}[k] N \sum_k r_k \ln r_k$, that reduces to Eqn. (11) only if both N and $\mathrm{E}[k] = \sum_k kp_k$ are constants. This is true when the constraint $\sum_k kp_k = \langle k \rangle$ is met. Moreover, the constraint $\sum_k kp_k = \langle k \rangle$ is equivalent to the normalization condition $\sum_k r_k = 1$ for the distribution r_k. Particularly, by using this normalization condition for maximizing S_r one obtains the uniform distribution,

$$r_k = \frac{1}{\Omega} \qquad \ldots (12)$$

equivalent to a microcanonical description of the statistical ensemble G_{nn} (supposed isolated).

A way to quantify the randomness of a degree distribution is to consider the nodes as members of a statistical ensemble. Thus, from a statistical-mechanical point of view, the uniform distribution $p_k = 1/\Omega$,

corresponding to a microcanonical description of the ensemble G_n, implies the most random evolution of the elements of G_n. Indeed, we may suppose that a given node V_j tends to establish a certain number of connections with other nodes of the graph, during its relaxation towards equilibrium. But, lacking any information, we do not know how many connections there will be among V_j and other nodes. A priori, the node V_j can assume any value of the connectivity, thus randomly choosing the number of its neighbour nodes. In other words, node V_j can belong to each of the Ω clusters of G_n with equal probability, just as one finds by maximizing entropy S_p only under the normalization condition $\sum_k p_k = 1$. At the same way, the uniform distribution $r_k = 1/\Omega$ implies the most random evolution of the elements of G_{nn}. Indeed, a priori, the node V_j may be linked to a neighbour that has any value of the connectivity, thus randomly choosing (k_j times) the connectivity of its neighbour nodes. In other words, each of the k_j copies of the node V_j can belong to each of the Ω clusters of G_{nn} with equal probability, just as we obtain by maximizing entropy S_r only subject to the normalization constraint $\sum_k r_k = 1$.

In conditions of maximum randomness, since each node of the graph belongs to both statistical ensembles, it would tend to simultaneously satisfy both distributions (3) and (12), which characterize the microcanonical equilibrium states of G_n and G_{nn}, respectively. But this is not possible as their probability distributions (p_k and r_k) are related, holding $r_k = kp_k/\mathrm{E}[k]$. Because of that, the two statistical ensembles cannot be considered isolated and they will never reach both uniform configurations (3) and (12) at the same time. Next, we explain under what hypotheses this behaviour may lead to a canonical equilibrium state, at witch the corresponding degree distributions of both nodes and nearest neighbours are scale-free.

Entropy Competition

Given a probability distribution, like the degree distribution p_k, its generating function is defined as,

$$G_p(z) = \sum_{k=0}^{\infty} p_k z^k \qquad \ldots (13)$$

which is absolutely convergent for all $|z| \leq 1$. Since the probability distribution p_k is assumed correctly normalized, one has that $G_p(1) = \sum_k p_k = 1$. Another important property is that the mean degree of the network is given by $G'_p(1) = \sum_k k p_k$. Given the generating function $G_p(z)$, it is easy to demonstrate [44] that the generating function $G_r(z)$ of the degree r_k of the nearest neighbours of a randomly chosen node is given by,

$$G_r(z) = \sum_{k=0}^{\infty} r_k z^k = \frac{zG'_p(z)}{G'_p(1)} \qquad \ldots (14)$$

that satisfies the normalization condition $\sum_k r_k = 1$, as $G_r(1) = 1$.

Particularly, with regard to the uniform degree distribution, given by Eqn. (3), one has that $G_p(z) = \sum_k z^k/\Omega$ and $G'_p(z) = \sum_k k z^{k-1}/\Omega$, from which, using Eqn. (14), it follows that $G_r(z) = \sum_k (k/\Lambda_1) z^k$, where with $\Lambda_1 = \sum_k k$ (the cut-off condition $0 \leq k \leq K$ ensures that Λ_1 is finite) we have denoted the normalization constant. Consequently, the distribution of the degree of the first neighbours, corresponding to the uniform degree distribution (3), is given by,

$$\rho_k = \frac{k}{\Lambda_1} \qquad \ldots (15)$$

At the same way, with regard to the degree distribution given by the following equation,

$$\pi_k = \frac{1}{k\Lambda_2} \qquad \ldots (16)$$

where $\Lambda_2 = \sum_k 1/k$ (Λ_2 is again finite for the cut-off condition) is the normalization constant, one has that $G_p(z) = \sum_k z^k/k\Lambda_2$ and $G'_p(z) = \sum_k z^{k-1}/\Lambda_2$, from which it follows that $G_r(z) = \sum_k z^k/\Omega$, again by using Eqn. (14). Thus the particular degree distribution π_k, defined by Eqn. (16), corresponds to the uniform distribution (12) of the degree of the first neighbours.

Both probability distributions p_k and r_k will never be uniform at the same time: each attempt, from the elements of G_n to assume the configuration that satisfies the uniform distribution (3), will upset the elements of G_{nn}, whose distribution will go away from that uniform, given by Eqn. (12), and *vice versa*. Thus, the relaxation of G_n towards its microcanonical equilibrium state (corresponding to the unconstrained maximum of S_p) will be constrained by the concomitant relaxation of G_{nn} towards its own microcanonical equilibrium (unconstrained maximum of S_r). Then, either system, coming in mutual competition, will tend to a canonical equilibrium state, at which the attempts to get uniform both distributions p_k and r_k mutually balance. Then, this conflictual behaviour can be assumed as the most responsible for the scale-free topology of the real-world complex networks, by assuming that the opposite tendency to the respective equilibrium states gives rise to the most probable distributions of both p_k and r_k. To this end, we utilized in [41] a condition on the probability distribution that

maximizes entropy S_r as a constraint in order to maximize the entropy S_p (or vice versa), in a way that we will explain shortly.

In Section III A we have introduced the information function $I(p_k) = -\ln p_k$, whose average value $E_p[I(p_k)] = -\sum_k p_k \ln p_k$ (entropy of message) measures how random the degree distribution of a complex network is. Now, let π_k be an a priori choice of the probability distribution p_k, the cross entropy is defined as the average value $E_p[I(\pi_k)] = -\sum_k p_k \ln \pi_k$ (on the probability space p_k) of the information related to the selected probability π_k. Cross-entropy function has been largely employed in linguistics [54] and in problems of data compression [55], and it is related to Kullback-Leibler divergence $D(p_k||\pi_k) = \sum_k p_k \ln(p_k/\pi_k)$ [52, 56], being $E_p[I(\pi_k)] = E_p[I(p_k)]+D(p_k||\pi_k)$. Keeping this in mind, let us find the most probable distribution of the elements of G_n. As said above, if G_n and G_{nn} are statistically independent, both systems would reach their own microcanonical equilibrium states. Particularly, the most random evolution of G_{nn} (supposed isolated) would lead to the uniform distribution (12), corresponding to the unconstrained maximum of S_r. Moreover, this uniform distribution corresponds to the particular distribution π_k, defined by Eqn. (16). Now, we think of the distribution π_k as the best choice of the degree distribution p_k, by assuming that the only available information on the evolution of G_n is given by the knowledge of the effects (π_k) that the most probable evolution of G_{nn} would produce on the ensemble G_n, if no other conditions are specified. So, in order to find the most probable distribution of the elements of G_n, it seems quite natural to introduce the information related to the distribution π_k as a constraint to maximize the entropy S_p. Such a constraint can be expressed in a natural way by making use of the cross entropy previously introduced. To this end, we impose that the mean value of the information related to the distribution π_k equals its expected value,

$$-\sum_{k=0}^{K} p_k \ln \pi_k = \langle I(\pi_k) \rangle \qquad \ldots (17)$$

From the point of view of statistical inference, we have hypothesized that the competitive behaviour between two ensembles represents the only information by means of which we can describe their statistical states. In other words, the least biased information that we can use to reduce the uncertainty-entropy of G_n (compared with its unconstrained maximum) is given by the knowledge of the condition of maximum uncertainty-entropy of G_{nn}, and vice versa.

Power-Law Degree Distribution

Given a graph, in the following we assure that there are no isolated vertices (i.e., $N_k(N) = 0$ for $k = 0$), and so in all summations in Eqns. (2), (11), and (17) k ranges from 1 to K. In [41] we defined the following constrained maximum entropy problem: let us find the maximum value of entropy S_p under the normalization constraint $\sum_k p_k = 1$, plus the following condition on the logarithmic mean of the connectivities,

$$\sum_{k=1}^{K} p_k \ln k = \langle \ln k \rangle_p \qquad \ldots (18)$$

that follows by substituting Eqn. (16) in the constraint (17). The pedix "p" to the symbol of expected value $\langle\ \rangle$ in Eqn. (18) denotes that the mean is calculated on the probability space p_k. From a purely mathematical point of view, the only difference between this problem and the one that has led to the Boltzmann distribution $p_k = e^{-\mu k}/Z$ discussed in Section III A, lies in substituting the condition $\sum_k k p_k = \langle k \rangle$ on the mean degree with the constraint (18) on the logarithmic mean $E[\ln k]_p$. The Lagrangian functional $L(p_k)$ assumes the following expression,

$$L(p_k) = -\sum_{k=1}^{K} p_k \ln p_k - \lambda \sum_{k=1}^{K} p_k + \lambda - \mu \sum_{k=1}^{K} p_k \ln k + \mu \langle \ln k \rangle_p \qquad \ldots (19)$$

where λ and μ are two Lagrange multipliers. Equating to zero the first derivatives of the Lagrangian, we obtain,

$$p_k = \frac{k^{-\mu}}{e^{1+\lambda}} = \frac{k^{-\gamma_p}}{Z_n} \qquad \ldots (20)$$

where we have made the change of notation $\gamma_p = \mu$ for later convenience, and with $Z_n = e^{1+\lambda}$ we have denoted the partition function. Substituting Eqn. (20) in the normalization constraint $\sum_k k p_k = \langle k \rangle$, one has that $Z_n = \zeta(\gamma_p)$, where

$$\zeta(\gamma_p) = \sum_{k=1}^{\infty} k^{-\gamma_p} \qquad \ldots (21)$$

is the Riemann zeta function, defined for all $k \geq 1$ and convergent if $\gamma_p > 1$. Thus, the (node) degree distribution assumes the following power-law form,

$$p_k = \frac{k^{-\gamma_p}}{\zeta(\gamma_p)} \qquad \ldots (22)$$

that stands for the canonical distribution of the elements of G_n. Moreover, by substituting Eqn. (22) in the constraint (18) we obtain,

$$-\frac{1}{\zeta(\gamma_p)}\frac{\partial\zeta(\gamma_p)}{\partial\gamma_p}=\langle\ln k\rangle_p \qquad \ldots (23)$$

that defines the values of the scaling exponent γ_p (see Figure 3) as a function of only the estimate of logarithmic mean of the connectivities, which is an intrinsic quantity of the whole network.

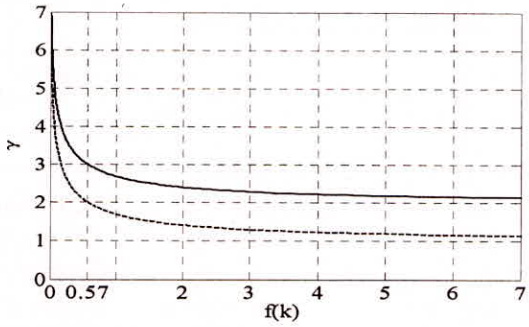

Fig. 3: Scaling exponent γ of the power-laws as a function of the topological features $f(k)$ of the network: the solid line represents γ_r as a function of $f(k) = \langle \ln k\rangle_r$, following Eqn. (18). Note that $\gamma_r<3$ approximately for $\langle \ln k\rangle_r>0.57$. The dashed line represents γ_p as a function of $f(k) = \langle \ln k\rangle_p$, according to Eqn. (13). In this case $\gamma_p<2$ approximately for $\langle \ln k\rangle_p>0.57$

By denoting with $\mathrm{Li}_n(z)$ the nth polilogarithm of z, the generating function of the degree distribution (22) may be written as,

$$G_p(z)=\frac{Li_{\gamma_p}(z)}{\zeta(\gamma_p)} \qquad \ldots (24)$$

from which, by using Eqn. (14), it follows that,

$$G_r(z)=\frac{Li_{\gamma_p-1}(z)}{\zeta(\gamma_p-1)} \qquad \ldots (25)$$

Consequently, the degree distribution r_k of the nearest neighbours of a randomly selected node is given by the following power-law,

$$r_k=\frac{k^{-(\gamma_p-1)}}{\zeta(\gamma_p-1)} \qquad \ldots (26)$$

convergent for all $\gamma_p>2$. Again in analogy with the statistical mechanics, the partition and Hamiltonian functions of the balanced ensemble G_n are given by $Z_n = \zeta(\gamma_p)$ and $H_n = \gamma_p \ln k$, respectively. Now, let us denote with T_p a generalized temperature of the ensemble G_n, for the above position we can assume $T_p \propto 1/\gamma_p$ and so, for high temperatures ($\gamma_p \to 0$), the pair of relations (22) and (26) becomes equal to the one defined by Eqns. (3) and (15). In these conditions, the canonical ensemble, whose weights are just defined by Eqn. (22), would coincide with the microcanonical one, defined by Eqn. (3).

Dual Problem

In the previous subsection we have explained the constrained maximum problem (that, in the following, we will denote with $S_p|S_r$) controlled by the Eqns. (2), (18) plus the normalization condition $\sum_k p_k = 1$. As showed in [41], this problem can be completely reversed. Indeed, each vertex of a graph is both node and nearest neighbour of other nodes, and so the role played by this element is interchangeable, belonging to both statistical ensembles G_n and G_{nn}. Then, in a complementary way to the one discussed in the previous subsection, we can maximize entropy S_r by using a constraint on the expected value of the information related to the distribution ρ_k (defined by Eqn. (15)), corresponding to the unconstrained maximum of entropy S_p. In this case, such a constraint can be expressed by means of cross entropy $-\sum_k r_k \ln \rho_k$, defined on the probability space r_k. Thus, in a similar way to that made for the method $S_p|S_r$, we introduce the dual problem $S_r|S_p$, given by the maximization of the entropy S_r under the normalization constraint $\sum_k r_k = 1$, plus the following condition on the logarithmic mean of the connectivities,

$$\sum_{k=1}^{K} r_k \ln k = \langle \ln k\rangle_r \qquad \ldots (27)$$

where the pedix "r" to the symbol of expected value denotes that the mean is calculated on the probability space r_k. The Lagrangian functional $L(r_k)$ is given by,

$$L(r_k)=-\sum_{k=1}^{K} r_k \ln r_k - \lambda\sum_{k=1}^{K} r_k + \lambda - \mu\sum_{k=1}^{K} r_k \ln k + \mu\langle \ln k\rangle_r \qquad \ldots (28)$$

with λ and μ Lagrange multipliers. Equating to zero the first derivatives of $L(r_k)$, we obtain,

$$r_k=\frac{k^{-\mu}}{e^{1+\lambda}}=\frac{k^{-(\gamma_r-1)}}{Z_{nn}} \qquad \ldots (29)$$

where we have made the change of notation $\gamma_r - 1 = \mu$, and with $Z_{nn} = e^{1+\lambda}$ we have denoted the partition function. Without going into details, proceeding as for the $S_p|S_r$ method, we obtain the following pair of power-laws,

$$p_k = \frac{k^{-\gamma_r}}{\zeta(\gamma_r)}, \text{ and } r_k = \frac{k^{-(\gamma_r - 1)}}{\zeta(\gamma_r - 1)} \quad \ldots (30)$$

but with different values of the scaling exponent γ_r, in this case defined by,

$$-\frac{1}{\zeta(\gamma_r - 1)} \frac{\partial \zeta(\gamma_r - 1)}{\partial(\gamma_r - 1)} = \langle \ln k \rangle_r \quad \ldots (31)$$

for all $\gamma_r > 2$ (see Figure 3). Also in this case, the scaling exponent is only a function of the topological features of the network, expressed by the estimate $\langle \ln k \rangle_r$ of the logarithmic mean of the connectivities. The latter of Eqns. (30) defines the canonical distribution of the elements of G_{nn}, for the which the partition and Hamiltonian functions are given by $Z_{nn} = \zeta(\gamma_r - 1)$ and $H_{nn} = (\gamma_r - 1)\ln k$, respectively. Moreover, by denoting with T_r a generalized temperature of the balanced ensemble G_{nn}, for the above position we can assume $T_r \propto 1/(\gamma_r - 1)$ and so, for high temperatures ($\gamma_r \to 1$), the two distributions in Eqns. (30) become equal to the ones defined by Eqns. (16) and (12), respectively. Thus, the canonical ensemble, whose weights are defined by the latter of Eqns. (30), would coincide with the microcanonical one, defined by Eqn. (12).

Both methods $S_p|S_r$ and $S_r|S_p$ describe two alternative evolutions of a network, leading to the canonical equilibrium states of G_n and G_{nn}, respectively. Particularly, in the case of $S_p|S_r$ method we think of G_n as a system, whereas we consider G_{nn} as a portion of surroundings. On the contrary, in the case of the $S_r|S_p$ method, G_{nn} is assumed as system, G_n, standing for the surroundings. Yet, in both cases we obtain the same pair of power-laws, although with different values of the scaling exponent. This is due to the substantial symmetry of two methods, as we can pass from one to another by simply swapping the roles played by the two entropy functions. This symmetric behaviour suggests that the scale-free topology of the real-world complex networks is only rooted in laws intrinsic to the system, independently of any external shorting established by the observer. Nevertheless, the different values of the scaling exponents involve a symmetry breaking responsible for the twofold evolution of the network, which can evolve in the ways described by the $S_p|S_r$ method as by the dual one, depending on the particular conditions of the surroundings. This symmetry breaking can be ascribed to the information "hidden" in the normalization condition $\sum_k r_k = 1$, used as a constraint in the $S_r|S_p$ method. Indeed, as noted in Section IV A, the constraint $\sum_k k p_k = \langle k \rangle$ on the mean degree of the network, that is needed to correctly define entropy S_r, implies the normalization condition $\sum_k r_k = 1$. The constraint $\sum_k k p_k = \langle k \rangle$ is instead absent in the set of equations that define the $S_p|S_r$ method. On the other hand, since the clusters of either statistical ensemble are identified by means of the different values of the connectivities, it is natural to consider the mean degree of the network in order to find the equilibrium configuration. Now we note that, for a power-law, the mean degree is finite for all values of the scaling exponent $\gamma > 2$, as we can write $E[k] = \zeta(\gamma - 1)/\zeta(\gamma)$. So, the method $S_r|S_p$, giving values of γ_r (see Eqn. (31)) that range from 2 to ∞, provides a power-law degree distribution with finite average degree. Since most of the real networks have values of the scaling exponent ranging from 2 to 3, the method $S_r|S_p$ is the most suitable to describe the topological features of real-world networks that have a finite value of the mean degree. Indeed, as shown in the diagram of Figure 3, for values of $\langle \ln k \rangle_r$ nearly greater than 0.57, γ_r is less than 3, as most frequently observed in nature.

On the other hand, the networks with infinite mean degree (i.e., with values of the scaling exponent between 2 and 1) may be more appropriately described by the method $S_p|S_r$, which does not impose any constraint on the average degree of the network. Indeed, this last method provides values of the scaling exponent γ_p theoretically greater than 1 (see Eqn. (23)), but that typically range from 1 to 2, as shown in Figure 3. In this way, the method $S_p|S_r$ also provides for the existence of those complex networks that have the lower values of the scaling exponent (i.e., with $\gamma < 2$).

In this work we showed how a power-law degree distribution can follow from a particular optimizing problem, since building a graph starting from a given degree sequence is a problem largely discussed by means of the CM, described in Section II A. Then, if we identify a particular mechanism generating a theoretical power-law degree distribution, such as the entropy competition, we should be confident that the same mechanism also determines the scale-free topology of the underlying graph. Nevertheless once achieved a scale-free degree distribution with scaling exponent values between $2 < \gamma \leq 3$ (i.e., $S_r|S_p$ method), we need to impose the structural cut-off to the degree distribution defined by the former of Eqns. (30) in order to build a simple graph using the CM. On the other hand, for power-law degree distributions with scaling exponent values between 1 and 2 (which dominate for the $S_p|S_r$ method) multiple edges and self-loops cannot be neglected by using the structural cut-off [57]. However, even in the range $1 < \gamma \leq 2$, results are interesting. In facts, in this case the CM generates multigraphs [57], which may be thought of as weighted networks [58], where the weight stands for

the number of multiple edges between two nodes, and the degree distribution should represent more correctly the distribution of expected weights of a given node.

Finally, by substituting the power-law degree distribution in the condition $Q(D) > 0$, Aiello *et al.* [46] have proved that the underlying graph has a giant component for all $\gamma < \gamma_0 \approx 3.4785$, where the critical value γ_0 is the root of $\zeta(\gamma - 2) - 2\zeta(\gamma - 1) > 0$. In our case, since both methods $S_p|S_r$ and $S_r|S_p$ give values of the scaling exponent typically belonging to the ranges $1 < \gamma_p < 2$ and $2 < \gamma_r < 3$, respectively, the above condition is met.

CONCLUSIONS

Real-world networks may be thought of as complex systems, whose whole structure is the result of the local interactions of a number of components, originating collective and a priori unexpected properties, such as the scale-free degree distribution, which affect the underlying dynamics in a fundamental way. An intriguing example of such a control between form and function is offered by the fluvial geomorphology. In facts, river networks display highly complex behaviours analogous to critical phenomena, including strong nonlinearities, thresholds, self-organization and fractal structures, whose recurrent patterns arising from optimal design exercise a dominant control over catchment-scale hydrological response.

The observation that scaling properties are common to many real networks, ranging from earth science to biology, suggests that the structure and the evolution of such systems are governed by simple and universal organizing principles, independent of the particular nature of the system components. In this sense, optimality criteria and statistical-mechanical tools have recognized to play a fundamental and unifying role in explaining the tendency of natural systems, ranging from tree-like structures to loop-dominated complex networks, to evolve toward an optimal and robust topology.

This paper has reviewed how the degree distributions that define three classes of complex networks can be derived by means of the maximum entropy principle, in accordance with the role played by the connectivity property of the nodes. Specifically, by ignoring the networked structure and thinking of the connectivity as an indefinite property of the nodes, an entropy function, similar to the Boltzmann-Gibbs one, can be defined. The maximization of such entropy, under a constraint on the expected value of the mean degree, leads to an exponential degree distribution, similar to the energy distribution in a gas system. On the other hand, by thinking of the connectivity as the number of the half-edges going out from a node, one can define a different entropy function, describing the statistical properties of hybrid elements, assembled by assigning to each node the number of its own half-edges. The maximization of this entropy function, again constrained by the expected value of the mean degree, leads to a Poissonian degree distribution. Finally, to investigate the origin of the power-law degree distribution, we have interpreted the connectivity not as a property of the nodes, but rather as a particular type of element that is the number of the first neighbours of a give node.

Due to this interpretation, one can define the statistical ensemble of the nearest neighbours, showing how it comes in conflict with the ensemble of the nodes. This is the key point of our approach, as the interwoven destinies of two statistical ensembles beget the mutually opposite tendency to respective equilibrium states, that we explained by means of the entropy competition concept. Based on this behaviour, the power-law degree distribution of the real-world complex networks has been interpreted as the most probable distribution that the elements (nodes and first neighbours) of an open complex system exhibit at a canonical equilibrium states, when both entropy functions try to reach their maxima, under mutual constraint.

Accordingly, the scale-free topology of real-world networks seems to emerge spontaneously as a result of a general law of nature, thus explaining the ubiquity of such a feature in a wide variety of complex systems.

REFERENCES

[1] Barabási, A.-L. and Albert R., "Emergence of Scaling in Random Networks." *Science* 286, 509 (1999); Barabási, A.-L., Albert, R. and Jeong, H. Mean-field theory for scale-free random networks. *Phys.* A 272, 173 (1999).

[2] Albert, R. and Barabási, A.-L., "Rev. Mod. Phys. Statistical Mechanics of Complex Networks." 74, 47 (2002); Dorogovtsev, S.N. and Mendes, J.F.F. "Evolution of networks." *Adv. Phys.* 51, 1079 (2002).

[3] Faloutsos, M., Faloutsos, P. and Faloutsos, C. (1999). "On Power-Law Relationships of the Internet Topology." *Computer Communication Review* 29, 251.

[4] Pastor-Satorras, R., Vázquez, A. and Vespignani, A., "Dynamical and Correlation Properties of the Internet." *Phys. Rev. Lett.* 87, 258701 (2001); Vázquez, A., Pastor-Satorras, R. and Vespignani, A. "Large-scale topological and dynamical properties of the Internet." *Phys. Rev.* E 65, 066130 (2002).

[5] Albert, R., Jeong, H. and Barabási, A.-L. (1999). "Diameter of the World-Wide Web." *Nature* 401, 130.

[6] Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N. and Barabási, A.-L., "The large-scale organization of metabolic networks." *Nature* 407, 651 (2000); Wagner, A. and Fell, D.A. "The small world inside large metabolic networks." *Proc. Royal Soc.* London B 268, 1803 (2001).

[7] Solé, R.V. and Montoya, J.M. "Complexity and Fragility in Ecological Networks." *Proc. Royal Soc.* London B 268, 2039 (2001); Montoya, J.M. and Solé, R.V., "Small World Patterns in Food Webs." *J. Theor. Biol.* 214, 405 (2002).

[8] Newman, M.E.J., "The structure of scientific collaboration networks." *Proc, Nat. Acad.* Sci U.S.A. 98, 404 (2001); Newman, M.E.J., Scientific collaboration networks. I. Network construction and fundamental results. Phys. Rev. E 64, 016131 (2001); Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. 64, 016132 (2001); Newman, M.E.J., Watts, D.J. and Strogatz, S.H., "Random graph models of social networks." *Proc. Nat. Acad. Sci.* U.S.A. 99, 2566 (2002).

[9] Amaral, L.A.N., Scala, A., Barthélémy, M. and Stanley, H.E. (2000). "Classes of small-world networks." *Proc. Natl. Acad. Sci.* U.S.A. 97, 11149.

[10] Tumminello, M., Lillo, F. and Mantenga, R.N. (2007). "Hierarchically nested factor model from multivariate data." *Europhys. Lett.* 78, 30006.

[11] Yamasaki, K., Gozolchiani, A. and Havlin, S., "Climate Networks around the Globe are Significantly Effected by El-Niño." Available at http://arxiv.org/abs/0804.1374

[12] Spanne, P. *et al.* (1994). "Synchrotron computed microtomography of porous media: topology and transports." *Phys. Rev. Lett.* 73, 2001.

[13] Valentini, L., Perugini, D. and Poli, G. (2007). "The "small-world" topology of rock fracture networks." *Physica A* 377, 323.

[14] Rinaldo, A., Banavar, J.R. and Maritan, A. (2006). "Trees, networks, and hydrology." *Water Resour. Res.* 42, 984.

[15] Molloy, M. and Reed, B., "A critical point for random graphs with a given sequence." *Random Structures and Algorithms* 6, 161 (1995); Molloy, M. and Reed, B. "The size of the giant component of a random graph with a given degree sequence." *Combinatorics, Probability and Computing* 7, 295 (1998).

[16] Liljeros, F., Edling, C.R., Amaral, L.A.N., Stanley, H.E. and Åberg, Y. (2001). "The web of human sexual contacts." *Nature*, 411, 907.

[17] Newman, M.E.J. (2005). "Power Laws, Pareto Distributions and Zipf's Law." *Contemporary Physics* 46, 323.

[18] Mandelbrot, B.B., Fractals: Form, Chance and Dimension. Freeman, W.H., New York (1977); The fractal Geometry of Nature. Freeman, W.H., New York (1983).

[19] Song, C., Havlin, S. and Make, H.A. (2005). "Self-similarity of complex networks." *Nature* 433, 392.

[20] Tarboton, D.G., Bras, R.L. and Rodríguez-Iturbe, I., "Fractal Nature of River Networks." *Water Resour. Res.* 24, 1317 (1988); Marani, A., Rigon, R. and Rinaldo, A. "A note on Fractal Channel Networks." *Water Resour. Res.* 27, 3041 (1991); Rodríguez-Iturbe, I., Marani, M., Rigon, R. and Rinaldo, A., "Self-organized river basin landscapes: Fractal and multifractal characteristics." *Water Resour. Res.* 30, 3531 (1994); Tarboton, D.G., "Fractal River Networks, Horton's Laws, and Tokunaga Cyclicity." *Hydrology* 187, 105 (1996).

[21] Horton, R.E., "Erosional Development of Streams and Their Drainage Basins: Hydrophysical Approach to Quantitative Morphology." *Bull. Geol. Soc. America* 56, 275 (1945); Strahler, A.N. "Quantitative Analysis of Watershed Geomorphology." *Trans. Am. Geophy. Union* 38, 913 (1957); Hack, J.T., "Studies of Longitudinal Stream Profiles in Virginia and Maryland." *US Geol. Survey Prof. Paper* 294-B, 45 (1957).

[22] Dodds, P.S. and Rothman, D.H. (1999). "Unified View of Scaling Laws in River Networks." *Physical Review* E 59, 4865.

[23] Maritan, A., Rinaldo, A., Rigon, Giacometti, R.A. and Rodríguez-Iturbe, I. (1996). "Scaling laws for river networks." *Physical Review* E 53, 1510.

[24] Sivapalan, M. (2005). "Pattern, Process and Function: Elements of a Unified Theory of Hydrology at the Catchment Scale." In *Encyclopedia of Hydrological Sciences*, p. 193, edited by Anderson, M.G., John Wiley & Sons.

[25] Beven, K.J. and Kirkby, M.J., "Towards a simple, physically based, variable contributing area model of catchment hydrology." *Int. Ass. Hydrol. Sci. Bull.* 24, 43 (1979); Rodríguez-Iturbe, I. and Valdés, J.B., "The Geomorphologic Structure of Hydrologic Response." *Water Resour. Res.* 15, 1409 (1979).

[26] Albert, R., Jeong, H. and Barabási, A.-B. (2000). "Error and attack tolerance of complex networks." *Nature* 406, 378.

[27] Jeong, H., Mason, S.P., Barabási, A.-L. and Oltvai, Z.N., "Lethality and centrality in protein networks." *Nature* 411, 41 (2001); Pastor-Satorras, R., Smith, E. and Solé, R.V., "Evolving protein interaction networks through gene duplication." *J. Theor. Biol.* 222, 199 (2003).

[28] Burda, Z., Correia, J.D. and Krzywicki, A., "Statistical ensemble of scale-free random graphs." *Phys. Rev. E* 64, 046118 (2001); Bogacz, L., Burda, Z. and Wacław, B. "Homogeneous complex networks." *Phys. A* 336, 587 (2006).

[29] Albert, R. and Barabási, A.-L. (2000). "Topology of evolving networks: local events and universality." *Phys. Rev. Lett.* 85, 5234.

[30] Dorogovtsev, S.N., Mendes, J.F.F. and Samukhin, A.N. (2000). "Structure of Growing networks: Exact Solution

of the Barabási-Albert's Model." *Phys. Rev. Lett.* 85, 4633.

[31] Krapivsky, P.L., Redner, S. and Leyvraz, F. (2000). "Connectivity of Growing Networks." *Phys. Rev. Lett.* 85, 4629.

[32] Scanlon, T.M., Taylor, K.K., Levin, S.A. and Rodríguez-Iturbe, I. "Positive feedbacks promote power-law clustering of Kalahari vegetation." *Nature* 449, 209 (2007); Kéfi, S., Rietker, M., Alados, C.L., Puedo, Y., Papanastasis V.P., ElAich, A. and de Ruiter, P.C. "Spatial vegetation patterns and imminent desertification in Mediterranean arid ecosystems." *Nature* 449, 213 (2007).

[33] i Cancho, R.F. and Solé, R.V., "Statistical Mechanics of Complex Networks." *Lecture Notes in Physics* (Springer, Berlin, 2003), Vol. 625, p. 114.

[34] Colizza, V., Banavar, J., Maritan, A. and Rinaldo, A. (2004). "Network Structures from Selection Principles." *Phys. Rev. Lett.* 92, 198701.

[35] Barthélemy, M. and Flammini, A. (2006). "Optimal Traffic Networks." *J. Stat. Mech.* L07002.

[36] Banavar, J.R., Maritan, A. and Rinaldo, A. (1999). "Size and form in efficient transportation networks." *Nature* 399, 130.

[37] West, G.B., Brown, J.H. and Enquist, B.J., "A general model for the origin of allometric scaling laws in biology." *Science* 276, 122 (1997); Enquist, B.J., Brown, J.H. and West, G.B., "Allometric scaling of plant energetics and population density." *Nature* 395, 163 (1998); Enquist, B.J., West, G.B., Charnov, E.L. and Brown, J.H., "Allometric scaling of production and life-history variation in vascular plants." *Nature* 401, 907 (1999).

[38] Rodríguez-Iturbe, I. and Rinaldo, A. (1997). "Fractal River Basins: Chance and Self-Organization." *Cambridge University Press*, New York.

[39] Park, J. and Newman, M.E.J. (2004). "The statistical mechanics of networks." *Phys. Rev. E 70*, 066117.

[40] Berg, J. and Lässig, M., "Correlated random networks." *Phys. Rev. Lett.* 89, 228701 (2002); Burda, Z. and Krzywicki, A., "Uncorrelated Random Networks." *Phys. Rev. E 67*, 046118 (2003); Dorogovtsev, S.N., Mendes, J.F.F. and Samukhin, A.N., "Principles of statistical mechanics of random networks." *Nucl. Phys. B 666*, 396 (2003); Palla, G., Derényi, I., Farkas, I. and Vicsek, T. "Statistical mechanics of topological phase transitions in networks." *Phys. Rev. E 69,* 046117 (2004); Fronczak, A., Fronczak, P. and Holyst, J.A., "Fluctuation-dissipation relations in complex networks." *Phys. Rev. E 73,* 016108 (2006); Bianconi, G., "The entropy of randomized network ensembles Europhys." *Lett. 81*, 28005 (2008).

[41] Sanchirico and Fiorentino. "Scale-free networks as entropy competition." *Submitted to Phys. Rev. E.*

[42] Erdös, P. and Rényi, A., "On random graphs." *Publicationes Mathematicae* 6, 290 (1959); Erdös, P. and Rényi, A., "On the evolution of random graphs." *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 5, 17 (1960).

[43] Bollobás, B. (1985). Random graphs, Academic Press, New York.

[44] Newman, M.E.J., Strogatz, S.H. and Watts, D.J. (2001). "Random graphs with arbitrary degree distribution and their applications." *Phys. Rev.* E 64, 026118.

[45] McKay, B.D. (1985). "Asymptotics for Symmetric 0–1 Matrices With Prescribed Row Sums." *Ars combinatorica* 19A, 15.

[46] Aiello, W., Chung, F. and Lu, L. (2000). "A Random Graph Model for Power law Graphs." *In Proceedings of the 32 and Annual ACM Symposium on Theory of Computing.*

[47] Boguñá, M., Pastor-Satorras, R. and Vespignani, A. (2004). "Cut-off and finite size effects in scale-free networks." *Eur. Phys. J. B* 38, 205.

[48] Dorogovtsev, S.N. and Mendes, J.F.F. and Samukhin, A.N. (2001). "Size-dependent degree distribution of scale-free growing network." *Phys. Rev. E 63*, 062101.

[49] Newman, M.E.J., "Assortative mixing in networks." *Phys. Rev. Lett.* 89, 208701 (2002); Maslov, S., Sneppen, K. and Zaliznyak, A., "Detection of Topological Patterns in Complex Networks: Correlation Profile of the Internet." *Phys. A 333,* 529 (2004).

[50] Shannon, C.E. (1948). "A Mathematical Theory of Communication." *Bell System Technical Journal 27*, 379 (1948); 27, 621.

[51] Jaynes, E.T. (1957). "Information theory and statistical mechanics." *Physical Review 106*, 620 (1957); 108, 171.

[52] Cover, T.M. and Thomas, J.A., Elements of Information Theory (John Wiley and Sons, New York, 1991).

[53] Boguñá, M. and Pastor-Satorras, R., "Class of correlated random networks with hidden variables." *Phys. Rev. E 68*, 036112 (2003); Catanzaro, M., Boguñá, M. and Pastor-Satorras, R., "Generation of uncorrelated random scale-free networks." *Phys. Rev. E 71*, 027103 (2004).

[54] Juola, P. (1998). In *Proceedings of New Methods in Language Processing 3*, Sidney.

[55] Baronchelli, A., Caglioti, E. and Loreto, V. (2005). *J. Stat. Mech.* P04002.

[56] Kullback, S. and Leibler, R.A. (1951). "On information and sufficiency." *Annals of Mathematical Statistics 22*, 79.

[57] Serrano, A. and Boguñá, M. (2005). "Weighted Configuration Model." *AIP Conference Proceedings* 776 (1), 101.

[58] Newman, M.E.J., "Analysis of Weighted networks." *Phys. Rev. E 70*, 056131 (2004); Barrat, A., Barthélémy, M., Pastor-Satorras, R. and Vespignani, A., "The architecture of complex weighted networks." Proc. *Natl. Acad. Sci.* U.S.A. 101, 3747 (2004).