Shannon Entropy and Degree Correlations in Complex Networks

SAMUEL JOHNSON, JOAQUÍN J. TORRES, J. MARRO, and MIGUEL A. MUÑOZ
Instituto Carlos I de Física Teórica y Computacional,
Departamento de Electromagnetismo y Física de la Materia,
Universidad de Granada, 18071 Granada.
SPAIN
samuel@onsager.ugr.es, jtorres@onsager.ugr.es, jmarro@ugr.es, mamunoz@onsager.ugr.es

Abstract: A wide range of empirical networks – whether biological, technological, information-related or linguistic – generically exhibit important degree-degree anticorrelations (i.e., they are *disassortative*), the only exceptions usually being social ones, which tend to be positively correlated (*assortative*). With a view to understanding where this universality originates, we obtain the Shannon entropy of a network and find that the partition of maximum entropy does not in general correspond to uncorrelated networks but, in the case of heterogeneous (scale-free) degree distributions, to a certain disassortativity. This approach not only gives a parsimonious explanation to a long-standing question, but also provides a neutral model against which to compare experimental data, and thus determine whether there are specific correlating mechanisms at work among the forces behind the evolution of a given real-world network.

Key–Words: Random graphs, assortativity, Shannon entropy.

1 Introduction

Over the past decade, the study of complex systems has been greatly facilitaded by an ever deeper understanding of their underlying structures – the topologies of their representations as graphs, or networks, in which nodes play the role of the (dynamical) elements and edges stand for some form of interaction [1]. Whether natural or artificial, complex networks have non-trivial topologies which are usually studied by analysing a variety of measures, such as the degree distribution, clustering, average paths, modularity, etc. [1, 2, 3] The mechanisms which lead to a particular structure and their relation to functional constraints are often not clear and constitute the subject of much debate [2, 3]. For instance, when nodes are endowed with some additional "property," a feature known as *mixing* or *assortativity* can arise, whereby edges are not placed between nodes completely at random, but depending in some way on the property in question. If similar nodes tend to wire together, the network is said to be *assortative* – while it is *disassortative* when nodes are preferentially connected to dissimilar neighbors. [4].

An interesting situation is when the property taken into account is the degree of each node – i.e., the number of neighboring nodes connected to it. It turns out that a high proportion of empirical networks – whether biological, technological, information-related or linguistic – are disassortatively arranged (high-degree nodes, or hubs, are preferentially linked to low-degree neighbors, and viceversa) while social networks are usually assortative. Such degree-degree correlations have important consequences for network characteristics such as connectedness and robustness [4].

However, while assortativity in social networks can be explained taking into account homophily [4] or modularity [5], the widespread prevalence and extent of disassortative mixing in most other networks remains somewhat mysterious. Maslov *et al.* found that the restriction of having at most one edge per pair of nodes induces some disassortative correlations in heterogeneous networks [6], and Park and Newman showed how this analogue of the Pauli exclusion principle leads to the edges following Fermi statistics [7] (see also [8]). However, this restriction is not sufficient to fully account for empirical data. In general, when one attempts to consider computationally all the networks with the same distribution as a given empirical one, the mean assortativity is not necessarily zero [9]. But since some "randomization" mechanisms induce positive correlations and others negative ones [10], it is not clear how the phase space can be properly sampled numerically.

In this paper, we show that there is a general reason, consistent with empirical data, for the "natural" mixing of most networks to be disassortative. Using an information-theory approach we find that the configuration which can be expected to come about in the absence of specific additional constraints turns out not to be, in general, uncorrelated. In fact, for highly heterogeneous degree distributions such as those of the ubiquitous scale-free networks, we show that the expected value of the mixing is usually disassortative: there are simply more possible disassortative configurations than assortative ones. This result provides a simple topological answer to a long-standing question. Let us caution that this does *not* imply that all scale-free networks are disassortative, but only that, in the absence of further information on the mechanisms behind their evolution, this is the neutral expectation.

The topology of a network is entirely described by its adjacency matrix $\hat{a}$, the element $\hat{a}_{ij}$ representing the number of edges linking node i to node j (for undirected networks, $\hat{a}$ is symmetric). Among all the possible microscopically distinguishable configurations a set of L edges can adopt when distributed among N nodes, it is often convenient to consider the set of configurations which have certain features in common – typically some macroscopic magnitude, like the degree distribution. Such a set of configurations defines an *ensemble*. In a seminal series of papers Bianconi has determined the partition functions of various ensembles of random networks and derived their statistical-mechanics entropy [12]. This allows the author to estimate the probability that a random network with certain constraints has of belonging to a particular ensemble, and thus assess the relative importance of different magnitudes and help discern the mechanisms responsible for a given real-world network. For instance, she shows that scale-free networks arise naturally when the total entropy is restricted to a small finite value. Here we take a similar approach: we obtain the Shannon information entropy encoded in the distribution of edges. As we shall see, both methods yield the same results [13, 14], but for our purposes the Shannon entropy is more tractable.

2 Entropy of networks

The Shannon entropy associated with a probability distribution p_m is $s = -\sum_m p_m \ln(p_m)$, where the sum extends over all possible outcomes m [14]. For a given pair of nodes (i, j), p_m can be considered to represent the probability of there being m edges between i and j. For simplicity, we will focus here on networks such that $\hat{a}_{ij}$ can only take values 0 or 1, although the method is applicable to any number of edges allowed. In this case, we have only two terms: $p_1 = \hat{\epsilon}_{ij}$ and $p_0 = 1 - \hat{\epsilon}_{ij}$, where $\hat{\epsilon}_{ij} \equiv E(\hat{a}_{ij})$ is the expected value of the element $\hat{a}_{ij}$ given that the network belongs to the ensemble of interest. The entropy associated with pair (i, j) is then $s_{ij} = -\left[\hat{\epsilon}_{ij}\ln(\hat{\epsilon}_{ij}) + (1-\hat{\epsilon}_{ij})\ln(1-\hat{\epsilon}_{ij})\right]$, while the total entropy of the network is $S = \sum_{ij}^N s_{ij}$:

$$S = -\sum_{ij}^{N}\left[\hat{\epsilon}_{ij}\ln(\hat{\epsilon}_{ij}) + (1-\hat{\epsilon}_{ij})\ln(1-\hat{\epsilon}_{ij})\right]. \quad (1)$$

Note that, since we have not imposed symmetry of the adjacency matrix, this expression is in general valid for directed networks. For undirected networks, however, the sum is only over $i \leq j$, with the consequent reduction in entropy. This expression for the Shannon entropy has been shown by Anand and Bianconi to be equivalent to the statistical mechanics entropy for the same ensemble [13].

For the sake of illustration, we will obtain the entropy of two different ensembles often used as models of random graphs: the fully random graph, or Erdős-Rényi ensemble, and the *configuration* ensemble with a scale-free degree distribution [2]. In this example, we assume the network to be sparse enough to expand the term $\ln(1-\hat{\epsilon}_{ij})$ in Eq. (1) and keep only linear terms. This reduces Eq. (1) to $S_{sparse} \simeq -\sum_{ij}^N \hat{\epsilon}_{ij}[\ln(\hat{\epsilon}_{ij}) - 1] + O(\hat{\epsilon}_{ij}^2)$. In the Erdős-Rényi ensemble, each of N nodes has an equal probability of receiving each of $\frac{1}{2}\langle k\rangle N$ undirected edges, where $\langle k\rangle$ is the mean degree. So, writing $\hat{\epsilon}_{ij}^{ER} = \langle k\rangle/N$, we have

$$S_{ER} = -\frac{1}{2}\langle k\rangle N\left[\ln\left(\frac{\langle k\rangle}{N}\right) - 1\right]. \quad (2)$$

The configuration ensemble, which imposes a given degree sequence $(k_1, ...k_N)$, is defined via the expected value of the adjacency matrix: $\hat{\epsilon}_{ij}^c = k_i k_j/(\langle k\rangle N)$ [2, 15]. This value leads to

$$S_c = \langle k\rangle N[\ln(\langle k\rangle N) + 1] - 2N\langle k \ln k\rangle,$$

where $\langle\cdot\rangle \equiv N^{-1}\sum_i(\cdot)$ stands for an average over nodes. The Internet, like many real-world networks, has a scale-free degree distribution, i.e. $p(k) \sim k^{-\gamma}$. For undirected networks of this kind, with an exponent $\gamma > 2$ and large N, the previous expression becomes

$$S_c(\gamma) = -\frac{\langle k\rangle N}{2}\left[\ln\left(\frac{\langle k\rangle}{N}\right) + 2\ln\left(\frac{\gamma-2}{\gamma-1}\right) - \frac{\gamma-4}{\gamma-2}\right]. \quad (3)$$

Fig. 1 displays the entropy per node for two ensembles form which the Intenet at the autonomous system level could have been taken, corresponding to various dates: an ER random network, Eq. (2), and a scale-free network with $\gamma = 2.3$, Eq. (3). The drop

ISSN: 1790-2769

ISBN: 978-960-474-189-2

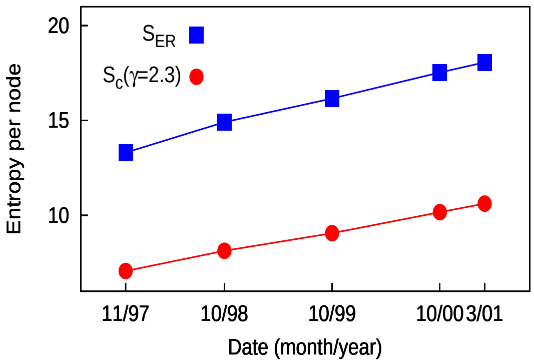

Figure 1: (Color online) Entropy per node of two ensembles which the Internet at the AS level beloged to at different dates [12]. Blue squares: Erdős-Rényi ensemble, from Eq. (2). Red circles: configuration ensemble, from Eq. (3) with $\gamma = 2.3$, respectively.

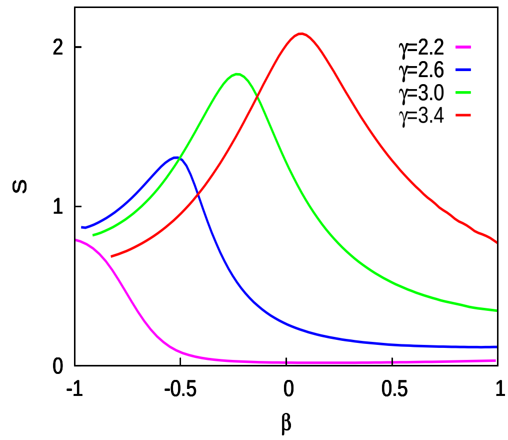

Figure 2: (Color online) Entropy of scale-free networks in the correlation ensemble against parameter β for various values of γ (increasing from bottom to top). $\langle k\rangle = 10$, $N = 10^4$, $C = 1$.

in entropy that comes about when the degree distribution is also considered highlights the unlikelihood of a scale-free distribution, and therefore the need for some specific mechanism (in this case, preferential attachement) to account for this feature.

3 Results

We will now go on to analyse the effect of degree-degree correlations on the entropy. In the configuration ensemble, the expected value of the mean degree of the neighbors of a given node is $k_{nn,i} = k_i^{-1}\sum_j \hat{\epsilon}^c_{ij} k_j = \langle k^2\rangle/\langle k\rangle$, which is independent of k_i. However, as mentioned above, real networks often display degree-degree correlations, with the result that $k_{nn,i} = k_{nn}(k_i)$. If $k_{nn}(k)$ increases (decreases) with k, the network is assortative (disassortative). A measure of this phenomenon is Pearson's correlation coefficient applied to the edges [2, 4, 3]: $r = ([k_l k_l'] - [k_l]^2)/([k_l^2] - [k_l]^2)$, where k_l and k_l' are the degrees of each of the two nodes belonging to edge l, and $[\cdot] \equiv (\langle k\rangle N)^{-1}\sum_l(\cdot)$ represents an average over edges. Writing $\sum_l(\cdot) = \sum_{ij}\hat{a}_{ij}(\cdot)$, r can be expressed in terms of averages over nodes:

$$r = \frac{\langle k\rangle\langle k^2 k_{nn}(k)\rangle - \langle k^2\rangle^2}{\langle k\rangle\langle k^3\rangle - \langle k^2\rangle^2}. \qquad (4)$$

The ensemble of all networks with a given degree sequence $(k_1, ...k_N)$ contains a subset for all members of which $k_{nn}(k)$ is constant (the configuration ensemble), but also subsets displaying other functions $k_{nn}(k)$. We can identify each one of these subsets (regions of phase space) with an expected adjacency matrix $\hat{\epsilon}$ which simultaneously satisfies the following conditions: **i**) $\sum_j k_j \hat{\epsilon}_{ij} = k_i k_{nn}(k_i)$, $\forall i$, and **ii**) $\sum_j \hat{\epsilon}_{ij} = k_i$, $\forall i$ (for consistency). An ansatz which fulfills these requirements is any matrix of the form

$$\hat{\epsilon}_{ij} = \frac{k_i k_j}{\langle k\rangle N} + \int d\nu \frac{f(\nu)}{N}\left[\frac{(k_i k_j)^\nu}{\langle k^\nu\rangle} - k_i^\nu - k_j^\nu + \langle k^\nu\rangle\right], \qquad (5)$$

where $\nu \in \mathbb{R}$ and the function $f(\nu)$ is in general arbitrary, although depending on the degree sequence it will here be restricted to values which maintain $\hat{\epsilon}_{ij} \in [0, 1]$, $\forall i, j$. This ansatz yields

$$k_{nn}(k) = \frac{\langle k^2\rangle}{\langle k\rangle} + \int d\nu f(\nu)\sigma_{\nu+1}\left[\frac{k^{\nu-1}}{\langle k^\nu\rangle} - \frac{1}{k}\right] \qquad (6)$$

(the first term being the result for the configuration ensemble), where $\sigma_{b+1} \equiv \langle k^{b+1}\rangle - \langle k\rangle\langle k^b\rangle$. In practice, one could adjust Eq. (6) to fit any given function $k_{nn}(k)$ and then wire up a network with the desired correlations: it suffices to throw random numbers according to Eq. (5) with $f(\nu)$ as obtained from the fit to Eq. (6) [16]. To prove the uniqueness of a matrix $\hat{\epsilon}$ obtained in this way (i.e., that it is the only one compatible with a given $k_{nn}(k)$) assume that there exists another valid matrix $\hat{\epsilon}' \neq \hat{\epsilon}$. Writting $\hat{\epsilon}'_{ij} - \hat{\epsilon}_{ij} \equiv h(k_i, k_j) = h_{ij}$, then **i**) implies that $\sum_j k_j h_{ij} = 0$, $\forall i$, while **ii**) means that $\sum_j h_{ij} = 0$, $\forall i$. It follows that $h_{ij} = 0$, $\forall j$.

In many empirical networks, $k_{nn}(k)$ has the form $k_{nn}(k) = A + Bk^\beta$, with $A, B > 0$ [3, 17] – the

ISSN: 1790-2769

ISBN: 978-960-474-189-2

mixing being assortative (disassortative) if β is positive (negative). Such a case is fitted by Eq. (6) if $f(\nu) = C[\delta(\nu - \beta - 1)\sigma_2/\sigma_{\beta+2} - \delta(\nu - 1)]$, with C a positive constant, since this choice yields

$$k_{nn}(k) = \frac{\langle k^2 \rangle}{\langle k \rangle} + C\sigma_2 \left[\frac{k^\beta}{\langle k^{\beta+1} \rangle} - \frac{1}{\langle k \rangle} \right]. \quad (7)$$

After plugging Eq. (7) into Eq. (4), one obtains:

$$r = \frac{C\sigma_2}{\langle k^{\beta+1} \rangle} \left(\frac{\langle k \rangle \langle k^{\beta+2} \rangle - \langle k^2 \rangle \langle k^{\beta+1} \rangle}{\langle k \rangle \langle k^3 \rangle - \langle k^2 \rangle^2} \right). \quad (8)$$

Inserting Eq. (5) in Eq. (1), we can calculate the entropy of correlated networks as a function of β and C – or, equivalently, as a function of r, by using Eq. (8). Particularizing for scale-free networks, then given $\langle k \rangle$, N and γ, there is always a certain combination of parameters β and C which maximizes the entropy; we will call these β^* and C^*. For $\gamma \lesssim 5/2$ this point corresponds to $C^* = 1$. For higher γ, the entropy can be slightly higher for larger C. However, for these values of γ, the assortativity r of the point of maximum entropy obtained with $C = 1$ differs very little from the one corresponding to β^* and C^* (data not shown). Therefore, for the sake of clarity but with very little loss of accuracy, in the following we will generically set $C = 1$ and vary only β in our search for the level of assortativity, r^*, that maximizes the entropy given $\langle k \rangle$, N and γ. Note that $C = 1$ corresponds to removing the linear term, proportional to $k_i k_j$, in Eq. (5), and leaving the leading non-linearity, $(k_i k_j)^{\beta+1}$, as the dominant one.

Fig. 2 displays the entropy curves for various scale-free networks, both as functions of β and of r: *depending on the value of γ, the point of maximum entropy can be either assortative or disassortative*. This can be seen more clearly in Fig. 3, where r^* is plotted against γ for scale-free networks with various mean degrees $\langle k \rangle$. The values obtained by Park and Newman [7] as those resulting from the one-edge-per-pair restriction are also shown for comparison: notice that whereas this effect alone cannot account for the Internet's correlations for any γ, entropy considerations would suffice if $\gamma \simeq 2.1$. Since most networks observed in the real world are highly heterogeneous, with exponents in the range $\gamma \in (2, 3)$, it is to be expected that these should display a certain disassortativity – the more so the lower γ and the higher $\langle k \rangle$. In Fig. 4 we test this prediction on a sample of empirical, scale-free networks cited in Newman's review [2] (p. 182). For each case, we found the value of r that maximizes S according to Eq. (1), after inserting Eq. (5) with the cited values of $\langle k \rangle$, N and γ. In this way, we obtained the expected assortativity for six

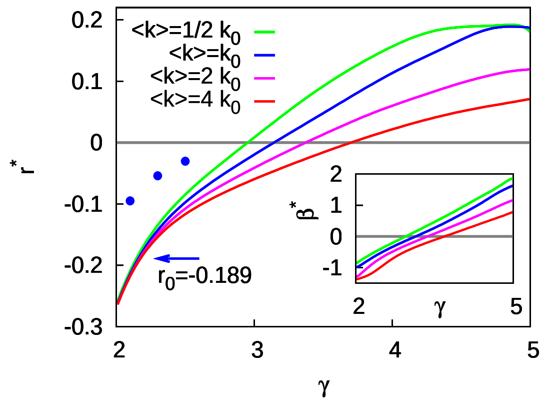

Figure 3: (Color online) Value of r at which the entropy is maximized, r^*, against γ, for random scale-free networks with $N = N_0 = 10697$ nodes and mean degrees $\langle k \rangle = \frac{1}{2}$, 1, 2 and 4 times $k_0 = 5.981$ (lines from top to bottom; N_0 and k_0 correspond to the values for the Internet at the AS level in 2001 [7], which had $r = r_0 = -0.189$). Circles are the values obtained in [7] as those expected soley due to the one-edge-per-pair restriction for k_0, N_0 and $\gamma = 2.1$, 2.3 and 2.5. Inset: r^* against N for networks with fixed $\langle k \rangle/N$ (same values as the main panel) and $\gamma = 2.5$; the arrow indicates $N = N_0$.

networks, representing: a peer-to-peer (P2P) network, metabolic reactions, the nd.edu domain, actor collaborations, protein interactions, and the Internet (see [2] and references therein). For the metabolic, Web domain and protein networks, *the values predicted are in excellent agreement with the measured ones*; therefore, no specific anticorrelating mechanisms need to be invoked to account for their disassortativity. In the other three cases, however, the predictions are not accurate, so there must be additional correlating mechanisms at work. Indeed, it is known that small routers tend to connect to large ones [17], so one would expect the Internet to be more disassortative than predicted, as is the case [18] – an effect that is less pronounced but still detectable in the more egalitarian P2P network. Finally, as is typical of social networks, the actor graph is significantly more assortative than predicted, probably due to the homophily mechanism whereby highly connected, big-name actors tend to work together [4].

4 Conclusions

We have shown how the ensemble of networks with a given degree sequence can be partitioned into re-

ISSN: 1790-2769

ISBN: 978-960-474-189-2

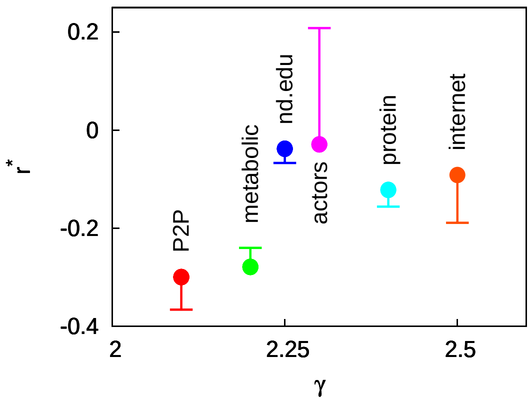

Figure 4: (Color online) Level of assortativity that maximizes the entropy, r^*, for various real-world, scale-free networks, as predicted theoretically by Eq. (1) against exponent γ. Bar ends show the empirical values.

gions of equally correlated networks and found, using an information-theory approach, that the largest (maximum entropy) region, for the case of scale-free networks, usually displays a certain disassortativity. Therefore, in the absence of knowledge regarding the specific evolutionary forces at work, this should be considered the most likely state. Given the accuracy with which our approach can predict the degree of assortativity of certain empirical networks *with no a priori information thereon*, we suggest this as a neutral model to decide whether or not particular experimental data require specific mechanisms to account for observed degree-degree correlations.

Acknowledgements: This work was supported by the Junta de Andalucía project FQM–01505, and by Spanish MICINN–FEDER project FIS2009–08451.

References:

[1] R. Albert A.-L. and Barabási, *Rev. Mod. Phys.* **74**, 47 (2002) S.N. Dorogovtsev and J.F.F. Mendes, *Evolution of networks: From biological nets to the Internet and WWW*, Oxford Univ, Press, 2003. R. Pastor-Satorras and A. Vespignani, *Evolution and structure of the Internet: A statistical physics approach*, Cambridge Univ. Press, Cambridge, 2004.

[2] M.E.J. Newman, *SIAM Reviews* **45**, 167 (2003).

[3] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Phys. Rep. **424**, 175 (2006).

[4] M.E.J. Newman, *Phys. Rev. Lett.*, **89**, 208701 (2002); *Phys. Rev. E* **67**, 026126 (2003).

[5] M.E.J. Newman and J. Park, *Phys. Rev. E* **68**, 036122 (2003).

[6] S. Maslov, K. Sneppen, and A. Zaliznyak, *Physica A* **333**, 529-540 (2004).

[7] J. Park and M.E.J. Newman, *Phys. Rev. E* **68**, 026112 (2003).

[8] A. Capocci and F. Colaiori, Phys. Rev. E. **74**, 026122 (2006).

[9] P. Holme and J. Zhao *Phys. Rev. E* **75**, 046111 (2007).

[10] I. Farkas, I. Derényi, G. Palla, and T. Vicsek, *Lect. Notes in Phys.* **650**, 163 (2004). S. Johnson, J. Marro, and J.J. Torres, *J. Stat. Mech., in press.* arXiv: 0905.3759.

[11] S. Johnson, J.J. Torres, J. Marro, and M.A. Muñoz, "The entropic origin of disassortativity in complex networks," *submitted to Phys. Rev. Lett.* (2010).

[12] G. Bianconi, *EPL* **81**, 28005 (2007); *Phys. Rev. E* **79**, 036114 (2009).

[13] K. Anand and G. Bianconi, *Phys. Rev. E* **80** 045102(R) (2009).

[14] E.T. Jaynes, *Physical Review* **106**, 620–630 (1957).

[15] S. Johnson, J. Marro, and J.J. Torres, *EPL* **83**, 46006 (2008).

[16] Although, as with the configuration ensemble, it is not always possible to wire a network according to a given $\hat{\epsilon}$.

[17] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, *Phys. Rev. Lett.* **87**, 258701 (2001).

[18] However, as Fig. 3 shows, if the Internet exponent were the $\gamma = 2.2 \pm 0.1$ reported in [17] rather than $\gamma = 2.5$, entropy would account more fully for these correlations.