PRAMANA
— journal of
physics
© Indian Academy of Sciences
Vol. 70, No. 6
June 2008
pp. 1135–1142

Most probable degree distribution at fixed structural entropy

GINESTRA BIANCONI
The Abdus Salam International Centre for Theoretical Physics, Strada Costiera 11, 34014 Trieste, Italy
E-mail: gbiancon@ictp.it

Abstract. The structural entropy is the entropy of the ensemble of uncorrelated networks with given degree sequence. Here we derive the most probable degree distribution emerging when we distribute stubs (or half-edges) randomly through the nodes of the network by keeping fixed the structural entropy. This degree distribution is found to decay as a Poisson distribution when the entropy is maximized and to have a power-law tail with an exponent $\gamma \to 2$ when the entropy is minimized.

Keywords. Complex networks; structural entropy.

PACS Nos 89.75.Hc; 89.75.Da; 89.75.Fb

1. Introduction

In the study of complex networks [1,2] the extension of the concept of entropy [3–6] has been recently shown to have many potential applications. It is indeed possible to define the *entropy of an ensemble* of networks as the normalized logarithm of the number of networks in a given ensemble [5]. Given a real network, different types of randomized network ensembles can be constructed, each one of them retaining structural information at different levels (degree sequence, degree correlations, community structure, distance in embedding space). The entropy of these network ensembles has been proposed as a measure of the relevance of the structural feature under consideration. Moreover, it is possible to define a *structural entropy*, i.e. the entropy of an uncorrelated network ensemble with given degree sequence. This entropy is the entropy of the configuration ensemble [7] and is the normalized logarithm of the number of possible networks spanned by the algorithm of reshuffling links proposed by Maslov and Sneppen [8]. In [5] it was shown that the structural entropy of power-law networks is an increasing function of the exponent $\gamma \in (2,3)$ of their degree distribution. Another definition regards the *entropy rate* of a diffusion process in a network, that carries information about the degree distribution and the degree correlations present in it [6]. In this paper we will first derive the expression [5] for the structural entropy. Secondly, we will show that the distribution which maximizes this entropy is extremely peaked around the average connectivity. Finally we will consider the degree distribution arising when we distribute stubs (or half-edges) randomly through the nodes of the network by keeping fixed the structural entropy. In this case the degree distribution that maximizes the entropy decays as a Poisson distribution. The degree distribution that minimizes the entropy instead has a tail which decays as a power-law with an exponent $\gamma \to 2$. This result indicates that scale-free degree distributions emerge naturally when considering networks ensemble with small structural entropy. The appearance of the power-law degree distribution reflects the tendency of social, technological and especially biological networks toward 'ordering'. This tendency is at work regardless of the mechanism which is driving their evolution that can be either a preferential attachment mechanism [9], or a 'hidden variables' mechanism [10–14] or some other statistical mechanics mechanisms [15,16].

2. The structural entropy

We will consider the ensemble of uncorrelated networks with N nodes and given degree sequence $\{k_1, \ldots, k_N\}$ with structural cut-off $k_i < K \ \forall i = 1, \ldots, N$. This ensemble is the configuration ensemble of networks with adjacency matrix $a_{i,j}$ such that $k_i = \sum_j a_{ij}$. The partition function of this network ensemble is given by

$$Z = \sum_{\{a_{ij}\}} \prod_i \delta\left(k_i - \sum_j a_{ij}\right). \tag{1}$$

We define the entropy per node Σ of the network ensemble as

$$\Sigma = \frac{1}{N} \ln Z. \tag{2}$$

Given (1) and expressing the delta's in the integral form with Lagrangian multipliers ω_i for every $i = 1, \ldots, N$ we get

$$Z = \int \mathcal{D}\omega \ \mathrm{e}^{-\sum_i \omega_i k_i} \prod_{i<j} (1 + \mathrm{e}^{\omega_i + \omega_j}), \tag{3}$$

where $\mathcal{D}\omega = \prod_i \mathrm{d}\omega_i/(2\pi)$. We solve this integral by saddle point equations. The entropy of this ensemble of networks can be approximated in the large network limit with

$$N\Sigma \simeq -\sum_i \omega_i^\star k_i + \sum_{i<j} \ln(1 + \mathrm{e}^{\omega_i^\star + \omega_j^\star}) \\ -\frac{1}{2}\sum_i \ln(2\pi\alpha_i) \tag{4}$$

with the Lagrangian multipliers ω_i satisfying the saddle point equations

$$k_i = \sum_{j \neq i} \frac{\mathrm{e}^{\omega_i^\star + \omega_j^\star}}{1 + \mathrm{e}^{\omega_i^\star + \omega_j^\star}}, \tag{5}$$

and the coefficients α_i defined as

$$\alpha_i = \sum_j \frac{\mathrm{e}^{\omega_i^\star+\omega_j^\star}}{\left(1+\mathrm{e}^{\omega_i^\star+\omega_j^\star}\right)^2}. \tag{6}$$

Given that the network has a structural cut-off, $k_i < \sqrt{\langle k\rangle N}$, $\forall i = 1\ldots, N$ we can approximate eq. (5) by $\mathrm{e}^{\omega_i} = k_i\sqrt{\langle k\rangle N}, \alpha_i = k_i$. In this limit the network is not correlated and the probability p_{ij} that a node i connects to a node j is given by $p_{ij} = k_i k_j/(\langle k\rangle N)$. Moreover $\omega_i^\star < 0$ and we can approximate the entropy as

$$\begin{aligned} N\Sigma \simeq & -\sum_i \ln[k_i/\sqrt{\langle k\rangle N}]k_i - \frac{1}{2}\sum_i \log(2\pi k_i) \\ & +\frac{1}{2}\sum_{ij}\frac{k_i k_j}{\langle k\rangle N} - \sum_{ij}\frac{1}{4}\frac{k_i^2 k_j^2}{(\langle k\rangle N)^2} + \cdots \\ = & -\sum_i (\ln k_i - 1)k_i - \frac{1}{2}\sum_i \ln(2\pi k_i) \\ & +\frac{1}{2}\langle k\rangle N[\ln(\langle k\rangle N) - 1] - \frac{1}{4}\left(\frac{\langle k^2\rangle}{\langle k\rangle}\right)^2 + \cdots \end{aligned} \tag{7}$$

which gives the following estimation for the number of networks in the ensemble:

$$\mathcal{N} \simeq \frac{(\langle k\rangle N)!!}{\prod_i k_i!}\exp\left[-\frac{1}{4}\left(\frac{\langle k^2\rangle}{\langle k\rangle}\right)^2\right] + \mathcal{O}(\log(N)) \tag{8}$$

with $\langle k\rangle N = 2L$, where L is the total number of links in the network and $(\cdots)!!$ indicates the double factorial.

3. Combinatorial interpretation of the entropy result

From combinatorial arguments we can derive an expression $\mathcal{N}_c$ for the number of networks with a given degree sequence which agree with the above extimate (8), i.e.

$$\log \mathcal{N}_c = \log \mathcal{N} + \mathcal{O}(\log N). \tag{9}$$

We want to show that by combinatorial arguments we can estimate the number of networks with given degree sequence to be

$$\mathcal{N}_c \propto \frac{(2L-1)!!\mathrm{e}^{-\frac{1}{4}(\frac{\langle k^2\rangle}{\langle k\rangle})^2}}{\prod_i k_i!}. \tag{10}$$

The total number of wiring of the links is given by $(2L-1)!!$. In fact let us construct a network from a set of N nodes, each node having a certain number of stubs departing from it. To wire the links as a first step we take a stub and we match it with one of the $2L-1$ other stubs of the network. Secondly, we match a new stub with one of the $2L-3$ remaining stubs. Repeating this procedure we get one out of $(2L-1)!!$ possible wirings of the links. This number includes also the wirings of the links which give rise to networks with double links. To estimate the number of such undesired wirings we assume that the network is random, i.e. the probability that a node with k_i stubs connects to a node with k_j stubs is a Poisson variable with average $k_i k_j/(\langle k\rangle N)$. In this hypothesis the probability Π that the networks do not contain double links is equal to [4]

$$\Pi = \prod_{i<j}\left(1+\frac{k_i k_j}{\langle k\rangle N}\right)\mathrm{e}^{-\frac{k_i k_j}{\langle k\rangle N}} \sim \mathrm{e}^{-\frac{1}{4}\left(\frac{\langle k^2\rangle}{\langle k\rangle}\right)^2}. \tag{11}$$

Finally in the expression (10) for $\mathcal{N}_c$ there is an additional term which takes into account the number of wirings of the links giving rise to equivalent networks. In graphs without double links this term is given by the number of possible permutation of the stubs at each node, i.e. $\prod_i k_i!$. We note here that a similar result was derived by mathematicians for the case in which the maximal connectivity $K < N^{1/3}$ [17] and an inequality was proved for the case $K > N^{1/3}$ [18]. Now we extend these results by statistical mechanics methods to uncorrelated networks with maximal connectivity $K < \sqrt{\langle k\rangle N}$.

4. Degree distribution maximizing the entropy of the network

The degree distribution N_k that maximizes the entropy when we take fixed the average degree and the number of nodes, is easily found by maximizing

$$N\Sigma = \log((2L)!!) - \sum_k N_k \log(k!) - \frac{1}{4}\sum_{k,k'} N_{k'} k N_k \frac{k^2 k'^2}{\langle k\rangle^2} \tag{12}$$

fixing the constraints

$$\sum_k N_k = N,$$

and

$$\sum_k k N_k = L. \tag{13}$$

We therefore have to differentiate $\mathcal{L}$ with respect to N_k with

$$\mathcal{L} = N\Sigma - \lambda\left(\sum_k k N_k - N\right) - \mu\left(\sum_k N_k - N\right). \tag{14}$$

We observe that

$$\frac{\partial \mathcal{L}}{\partial N_k} = -\log(k!) - \frac{k^2\langle k^2\rangle}{2\langle k\rangle} - \lambda k - \mu. \tag{15}$$

This derivative has a maximum $k = k^\star(\lambda, \mu)$ in which the derivative can be either positive or zero depending on the value of the Lagrangian multipliers λ, μ. Imposing the constraints (13) the resulting degree distribution is peaked at $k^\star = 2L/N$ with a distribution that is exactly $N_k = N\delta\left(k - \frac{L}{N}\right)$ if the ratio L/N is an integer. In the case where L/N is not an integer the degree distribution is minimally spread around two peaks at $[L/N]$ ($[\cdot]$ indicating the integer part) and $[L/N] + 1$.

5. Natural degree distribution at fixed structural entropy

We want to find the most probable degree distribution that we can obtain by distributing $2L$ stubs to N nodes in such a way that the network ensemble with that degree distribution has fixed structural entropy.

Proceeding as in the standard statistical mechanics, we define a normalized partition function $\mathcal{Z}$ of the network as the sum over all degree distributions, with given energy

$$\mathcal{Z} = \frac{1}{C}\sum_{\{N_k\}}{}' \mathcal{N}_{\{N_k\}} \mathrm{e}^{\beta N \Sigma(\{N_k\})} \tag{16}$$

with $C = (2L)! \exp[\beta(2L)!!]$.

The role of the parameter β is to fix the average value of the entropy Σ. When $\beta \to \infty$ the entropy Σ is maximized and when $\beta \to \beta_{\min}$ the entropy Σ is minimized. In formula (16) for the partition function $\mathcal{Z}$, we have introduced the factor $\mathcal{N}_{\{N_k\}}$ indicating the number of ways in which we can distribute $2L$ indistinguishable stubs obtaining the $\{N_k\}$ degree distribution. This term is given by the multinomial factor

$$\mathcal{N}_{\{N_k\}} = \frac{(2L)!}{\prod_k (kN_k)!}. \tag{17}$$

In fact we have to assign kN_k stubs to nodes of degree k.

In eq. (16) the sum $\sum'$ over the $\{N_k\}$ distributions is extended only to $\{N_k\}$ for which the total number of nodes N and the total number of links L in the network is fixed, i.e.

$$\sum_k N_k = N$$
$$\sum_k kN_k = 2L. \tag{18}$$

To enforce these conditions we introduce in (16) the delta functions in the integral form providing the expression

$$\begin{aligned}
\mathcal{Z} &= \int \mathrm{d}S \int \frac{\mathrm{d}\lambda}{2\pi} \int \frac{\mathrm{d}\mu}{2\pi} \int \frac{\mathrm{d}\nu}{2\pi} \\
&\quad \times \sum_{\{N_k\}} \exp\left[-\beta \sum_k N_k \log k! - \frac{\beta}{4}\left(\frac{S}{\langle k\rangle}\right)^2 - \sum_k \log[(kN_k)!] \right. \\
&\quad \left. -i\lambda\left(2L - \sum_k N_k k\right) - i\mu\left(N - \sum_k N_k\right) - i\nu\left(NS - \sum_k k^2 N_k\right)\right] \\
&= \int \mathrm{d}S \int \frac{\mathrm{d}\lambda}{2\pi} \int \frac{\mathrm{d}\mu}{2\pi} \int \frac{\mathrm{d}\nu}{2\pi} \\
&\quad \times \exp\left[-i\lambda 2L - i\mu N - i\nu NS - \frac{\beta}{2}\left(\frac{S}{\langle k\rangle}\right)^2 + \sum_k \log G_k(\lambda,\mu,\nu)\right] \\
&= \int \mathrm{d}S \int \frac{\mathrm{d}\lambda}{2\pi} \int \frac{\mathrm{d}\mu}{2\pi} \int \frac{\mathrm{d}\nu}{2\pi} \exp[Nf(\lambda,\mu,\nu,S)], \qquad (19)
\end{aligned}$$

where

$$G_k(\lambda,\mu,\nu) = \sum_{N_k} \frac{1}{(kN_k)!} \exp\left\{kN_k\left[i\lambda + i\frac{\mu}{k} + i\nu k - \frac{\beta}{k}\log(k!)\right]\right\}. \qquad (20)$$

Assuming that the sum over all N_k can be approximated by the sum over all $L_k = kN_k = 1, 2, \ldots, \infty$ we get $\log G_k(\lambda,\mu,\nu) = \exp\left[i\lambda + i\mu/k - \frac{\beta}{k}\log(k!) + i\nu k\right]$ and

$$\begin{aligned}
f(\lambda,\mu,\nu,S) &= -i\langle k\rangle\lambda - i\mu - i\nu S - \frac{\beta}{4}\left(\frac{S}{\langle k\rangle}\right)^2 \\
&\qquad + \frac{1}{N}\sum_k \mathrm{e}^{i\lambda + i\mu/k - \frac{\beta}{k}\log(k!) + i\nu k}, \qquad (21)
\end{aligned}$$

where $\langle k\rangle = 2L/N$ indicates the average degree of the network. By evaluating (19) at the saddle point, deriving the argument of the exponential with respect to λ and ν, we obtain

$$\begin{aligned}
1 &= \frac{1}{N}\sum_k \frac{1}{k}\mathrm{e}^{i\lambda + i\mu/k - \frac{\beta}{k}\log(k!)) + i\nu k}, \\
\langle k\rangle &= \frac{1}{N}\sum_k \mathrm{e}^{i\lambda + i\mu/k - \frac{\beta}{k}\log(k!)) + i\nu k}. \qquad (22)
\end{aligned}$$

With

$$S = \frac{1}{N}\sum_k k^2 \mathrm{e}^{i\lambda + i\mu/k - \frac{\beta}{k}\log(k!) + i\nu k} \qquad (23)$$

the value of ν to $\nu = 0$ is fixed. These equations always have a solution for sparse networks with $L = \mathcal{O}(N)$ provided $\beta > 1$ and $\langle k\rangle > 1$. The marginal probability that $L_k = kN_k$ is given by

$$P(L_k = kN_k) = \frac{1}{(kN_k)!} \mathrm{e}^{-\beta N_k \log(k!)} \frac{\mathcal{Z}_k(L, kN_k, N)}{\mathcal{Z}(L)}, \tag{24}$$

with

$$\mathcal{Z}_k(L, \ell, N) = \int \mathrm{d}S \int \frac{\mathrm{d}\lambda}{2\pi} \int \frac{\mathrm{d}\mu}{2\pi} \int \frac{\mathrm{d}\nu}{2\pi} \exp[N f_k(\lambda, \mu, \nu, S, \ell)] \tag{25}$$

and

$$\begin{aligned} f_k(\lambda, \mu, \nu, \ell) = &-i(\langle k\rangle - \ell/N)\lambda - i\mu(1 - \ell/(kN)) - i\nu(S - k\ell/N) \\ &- \frac{\beta}{4}\left(\frac{S^2}{\langle k\rangle}\right)^2 + \frac{1}{N}\sum_{s\neq k} \exp\left[i\lambda + i\mu/s - \frac{\beta}{s}\log(s!) + i\nu s\right]. \end{aligned}$$

If we develop (24) for $\ell \ll L$ and use the Stirling approximation for factorials, we get that each variable L_k is a Poisson variable with mean $\langle L_k\rangle$ satisfying

$$\frac{\langle L_k\rangle}{k} = \langle N_k\rangle \simeq k^{-\beta-1}\mathrm{e}^{i\lambda+\beta+i\mu/k}, \tag{26}$$

where we assumed that the minimal connectivity of the network is $k > 0$. The average $\langle N_k\rangle$ is a power-law distribution with a lower effective cut-off μ fixing and $\langle k\rangle$, a Lagrangian parameter λ fixing the normalization constant and β fixing the structural entropy. The distribution of $P(N_k)$ is finally

$$P(N_k) = \frac{k}{(kN_k)!}\mathrm{e}^{-\beta N_k \log(k!) + i\lambda k N_k + i\mu N_k}. \tag{27}$$

In the limit $\beta \to \infty$ $\langle N_k\rangle$ is extremely peaked on the average degree $k \simeq k^\star = \mathcal{O}(\langle k\rangle)$ of the network and the degree distribution N_k decays at large value of N_k as a Poisson distribution, i.e.

$$P(N_k) \simeq \frac{1}{(kN_k)!}\mathrm{e}^{N_k[-\beta\log(k^\star!) + i\lambda k^\star + i\mu]}. \tag{28}$$

When the value of the entropy is minimal, $\beta \to 1$ the degree distribution (28) has a large tail with an exponent $\gamma \to 2$.

6. Conclusions

Our results show that given a network with N nodes and L links, the best approximation to a regular network which is possible to construct maximizes the structural entropy. Moreover, we derive the degree distributions of networks with given value of the entropy when we randomly distribute stubs through the nodes. For small values of the entropy this distribution develops a power-law tail and exponent γ such that $\gamma \to 2$ when the entropy is minimal. On the contrary for high value of the entropy the distribution decays as a Poisson distribution.

Acknowledgments

This work was supported by IST STREP GENNETEC contract No. 034952.

References

[1] S N Dorogovtsev, A Goltsev and J F F Mendes, arXiv:0705.0010 [cond-mat] (2007)
[2] S Boccaletti, V Latora, Y Moreno, M Chavez and D U Hwang, *Phys. Rep.* **424**, 175 (2006)
[3] L Bogacz, Z Burda and B Waclaw, *Physica* **A366**, 587 (2006)
[4] G Bianconi, *Chaos* **17**, 026114 (2007)
[5] G Bianconi, *Europhys. Lett.* **81**, 28005 (2008)
[6] J Gomez-Gardenes and V Latora, arXiv:0712.0278 [cond-mat]
[7] M Molloy and B A Reed, *Rand. Stru. Alg.* **6**, 161 (1995)
[8] S Maslov and K Sneppen, *Science* **296**, 910 (2002)
[9] A-L Barabási and R Albert, *Science* **286**, 509 (1999)
[10] J Park and M E J Newman, *Phys. Rev.* **E70**, 066146 (2004)
[11] B Sodeberg, *Phys. Rev.* **E66**, 066121 (2002)
[12] F Chung and L Lu, *Proc. Natl. Acad. Sci.* **100**, 6313 (2002)
[13] G Caldarelli, A Capocci, P De Los Rios and M A Muñoz, *Phys. Rev. Lett.* **85**, 5468 (2002)
[14] M Boguñá and R Pastor-Satorras, *Phys. Rev.* **E68**, 036112 (2003)
[15] Z Burda, J D Correia and A Krzywicki, *Phys. Rev.* **E64**, 046118 (2001)
[16] S N Dorogovstev, J F F Mendes and A N Samukhin, *Nucl. Phys.* **B666**, 396 (2003)
[17] E Bender and E Rodney Canfield, *J. Combin. Theory Ser.* **A24**, 296 (1978)
[18] B D McKay, *Ars Combin.* **A19**, 15 (1985)