PHYSICAL REVIEW E **80**, 045102(R) (2009)

Entropy measures for networks: Toward an information theory of complex topologies

Kartik Anand[1] and Ginestra Bianconi[2]
[1]*The Abdus Salam International Center for Theoretical Physics, Strada Costiera 11, 34014 Trieste, Italy*
[2]*Department of Physics, Northeastern University, Boston, Massachusetts 02115, USA*
(Received 28 July 2009; published 13 October 2009)

The quantification of the complexity of networks is, today, a fundamental problem in the physics of complex systems. A possible roadmap to solve the problem is via extending key concepts of information theory to networks. In this Rapid Communication we propose how to define the Shannon entropy of a network ensemble and how it relates to the Gibbs and von Neumann entropies of network ensembles. The quantities we introduce here will play a crucial role for the formulation of null models of networks through maximum-entropy arguments and will contribute to inference problems emerging in the field of complex networks.

DOI: 10.1103/PhysRevE.80.045102 PACS number(s): 89.75.Hc, 89.75.Fb, 89.75.Da

I. INTRODUCTION

Complex networks [1–4] are found to characterize the underlying structure of many biological, social, and technological systems. Following ten years of active research in the field of complex networks, the state of the art includes a deep understanding of their evolution [1], an unveiling of the rich interplay between network topology and dynamics [3], and a description of networks through structural characteristics [2,4]. Nevertheless, we still lack the means to quantify, *how complex is a complex network*. In order to answer this question we need a new theory of information of complex networks. This theory will contribute to solving many challenging inference problems in the field [4–6]. By providing an evaluation of the information encoded in complex networks, this will resolve one of the outstanding problems in the statistical mechanics of complex systems.

In information theory [7] entropy measures play a key role. In fact, it is well known that the Shannon entropy and the von Neumann entropy are related to the information present in classical and quantum systems, respectively. Moreover, the aforementioned measures also have statistical mechanics interpretations. Traditionally, in statistical mechanics, for configurations drawn from canonical ensembles, the Shannon entropy corresponds to the entropy for classical systems, while the von Neumann entropy provides the statistical description of quantum systems.

In the context of complex networks a number of different entropy measures have been introduced [5,8–13]. In Ref. [9] the Gibbs entropy per node, in a network of N nodes, denoted by Σ, was introduced for microcanonical network ensembles following a statistical mechanics paradigm. Microcanonical network ensembles are defined as those networks that satisfy a given set of constraints. Examples of some popular constraints include fixed number of links per node, given degree sequence, and community structure. The Gibbs entropy of these ensembles is given by

$$\Sigma = \frac{1}{N}\log \mathcal{N}, \tag{1}$$

where $\mathcal{N}$ indicates the cardinality of the ensemble, i.e., the total number of networks in the ensemble. As demonstrated further in [9] the statistical mechanics formalism enables us to develop canonical network ensembles where the structural constraints under consideration are satisfied, *on average*. In classical statistical mechanics the microcanonical ensemble is formed by configurations having constant energy E, while the canonical ensemble is formed by configurations having constant average energy $\langle E\rangle$. By analogy, in the theory of random graphs the $G(N,L)$ graph ensemble is formed by networks of N nodes with a constant total number of links L. In the conjugated-canonical $G(N,p)$ ensemble, however, the total number of links is Poisson distributed with average $\langle L\rangle = p(N-1)$. This construction of microcanonical and conjugate-canonical ensemble can be further generalized [9] to network ensembles with more elaborate sets of constraints. For example, we can define microcanonical network ensembles with the given degree sequence $\{\kappa_i\}$ and canonical network ensembles (based on hidden variables [14,15]) in which each node i has k_i links, which is Poisson distributed with average $\langle k_i\rangle = \kappa_i$.

In this Rapid Communication we show for this statistical mechanics framework of networks, first, that the entropy of canonical network ensembles is related to the Shannon entropy and, second, that canonical network ensembles satisfy a principle of maximal Shannon entropy. Moreover we will study to what extent canonical and microcanonical network ensembles are equivalent. Finally we will discuss the relation between the Shannon entropy of a canonical network ensemble, S, and the recent definition of von Neumann entropy of networks, S_{VN}, recently introduced in Ref. [12] of interest in the field of quantum gravity [13].

II. GIBBS ENTROPY OF A MICROCANONICAL NETWORK ENSEMBLE

Microcanonical network ensemble are formed by network satisfying a given number of constraints. Following the lines of reasoning provided in [9], on specifying the full set of constraints and number of nodes N in the networks, one may introduce a partition function Z for the ensemble. This partition function counts the number of networks, defined by their adjacency matrices $\{a_{ij}\}$, that simultaneously satisfy all the constraints under consideration. The adjacency matrix describes an undirected network, i.e., $a_{ij} = a_{ji}$, where each element takes some positive integer values, $a_{ij} \in \alpha$, where $\alpha \subset \mathbb{N}$ that indicates the weight of a link between nodes i and j. For simple (connectivity) networks we take $a_{ij} \in \{0,1\}$ while for weighted networks $a_{ij} \in \mathbb{N}$.

TABLE I. Maximum-entropy networks ensembles with given set of constraints. The community of each node is associated with a Potts variable q_i. The distance of the nodes is binned and indicated by a discrete variable $d_{ij}=d$. The hidden variables of each ensembles $\{\theta_i\}$, $W(q,q')$, $W(d)$, $\{\alpha_i\}$, and $\{f_{ij},g_{ij}\}$ are fixed by respective conditions specified in the table.

Ensembles	Probabilities $p_{ij}/(1-p_{ij})$	Conditions	
Given expected number of links L	$p/(1-p)$	$pN(N-1)/2=L$	
Given expected community structure $\{A_{q,q'}\}$	$W(q_i,q_j)$		$A(q,q')\|_{q\neq q'}=\sum_{ij} p_{ij}\delta_{q_i,q}\delta_{q_j,q'}$
			$A(q,q)=\sum_{i<j} p_{ij}\delta_{q_i,q}\delta_{q_j,q}$
Given expected degree sequence $\{\kappa_i\}$	$\theta_i\theta_j$	$\kappa_i=\sum_j p_{ij}$	
Given expected degree sequence $\{\kappa_i\}$ community structure $\{A(q,q')\}$	$\theta_i\theta_j W(q_i,q_j)$	$\kappa_i=\sum_j p_{ij}$	$A(q,q')\|_{q\neq q'}=\sum_{ij} p_{ij}\delta_{q_i,q}\delta_{q_j,q'}$
			$A(q,q)=\sum_{i<j} p_{ij}\delta_{q_i,q}\delta_{q_j,q}$
Given expected degree sequence $\{\kappa_i\}$ and number of link at distance $d,B(d)$	$\theta_i\theta_j W(d_{ij})$	$\kappa_i=\sum_j p_{ij}$	$B(d)=\sum_{ij} p_{ij}\delta_{d_{ij},d}$
Given expected degree sequence $\{\kappa_i\}$ and number of triangles for each node $\{T_i\}$	$\theta_i\theta_j e^{f_{ij}(\alpha_i+\alpha_j)+g_{ij}}$	$\kappa_i=\sum_j p_{ij}$	$f_{ij}=\sum_k p_{ik}p_{kj}$
		$T_i=\sum_{jk} p_{ij}p_{jk}p_{ki}$	$g_{ij}=\sum_k p_{ik}\alpha_k p_{kj}$

Thus, we have

$$Z=\sum_{\{a_{ij}\}}\prod_k \delta(\text{constraint}_k(\{a_{ij}\}))\exp\left(-\sum_{i<j}\sum_\alpha h_{ij}(\alpha)\,\delta_{a_{ij},\alpha}\right), \tag{2}$$

where the fields $h_{ij}(\alpha)$ play the usual role of auxiliary fields in statistical mechanics. Finally the Gibbs entropy Σ, defined by Eq. (1), and the probability $\pi_{ij}(\alpha)$ of having a link between nodes i and j, with weight α, are given by

$$N\Sigma = \log Z|_{h_{ij}(\alpha)=0\forall(i,j,\alpha)},$$

$$\pi_{ij}(\alpha)=\frac{\partial \log Z}{\partial h_{ij}(\alpha)}. \tag{3}$$

III. ENTROPY OF A CANONICAL NETWORK ENSEMBLES

The canonical network ensemble can be built starting from the marginal distribution $\pi_{ij}(\alpha)$ given by Eq. (3). For a network of N nodes, for each pair of nodes, (i,j), one draws a link of weight α with probability $\pi_{ij}(\alpha)$. The probability Π of the canonical undirected network ensemble defined by its adjacency matrix $\{a_{ij}\}$ is therefore given by

$$\Pi=\prod_{i<j}\pi_{ij}(a_{ij}), \tag{4}$$

for which the logarithmic-likelihood function is given by

$$\mathcal{L}=-\sum_{i<j}\log\,\pi_{ij}(a_{ij}). \tag{5}$$

The entropy of a canonical ensemble is the logarithm of the number of typical networks in the ensembles and is given by

$$S=\langle\mathcal{L}\rangle_\Pi=-\sum_{i<j}\sum_\alpha \pi_{ij}(\alpha)\log\,\pi_{ij}(\alpha), \tag{6}$$

which takes exactly the form of a Shannon entropy. We will therefore call this quantity the Shannon entropy of a network ensemble. In particular, for the case of a simple undirected network, where $\alpha\in\{0,1\}$, we have

$$S=-\sum_{i<j} p_{ij}\log p_{ij}-\sum_{i<j}(1-p_{ij})\log(1-p_{ij}), \tag{7}$$

where $p_{ij}=\pi_{ij}(1)$ is the probability of having a link between nodes i and j.

Maximizing the Shannon entropy of the network subjected to different types of constraints gives rise to maximum-entropy ensembles and generalizing the maximum-likelihood arguments of [16]. In the following we will consider few examples of such constraints for the cases of simple undirected networks.

Fixing the total number of expected links, $\sum_{ij} p_{ij}=L$, the maximum-entropy ensemble is $G(N,\{p_{ij}\})$, with $p_{ij}=p=L/[N(N-1)/2]$. Alternatively, if we constrain the expected degree of each node i, i.e., $\kappa_i=\sum_j p_{ij}$, the probabilities in the maximum-entropy ensemble take the form $p_{ij}=\theta_i\theta_j/(1+\theta_i\theta_j)$, where θ_i are hidden variables fixed by the constraints. This ensemble is the canonical conjugated to the microcanonical ensemble of networks with fixed degree sequence

$\{\kappa_i\}$. In Table I we generalize this construction and report the form of maximum-entropy network ensembles satisfying a different sets of constraints. We leave to the reader the construction of maximum-entropy weighted network ensembles related to the canonical ensembles discussed in Refs. [9,17]. The marginal probability for the microcanonical and conjugated canonical ensembles are equal by definition, but in order to prove the equivalence between the two ensembles also the entropy per node Σ and S/N must be equal in the thermodynamic limit.

IV. COMPARISON BETWEEN THE ENTROPIES OF THE $G(N,L)$ AND THE $G(N,p)$ ENSEMBLES

We study first the relation between the Gibbs entropy Σ and the Shannon entropy per node for random graphs defined for the $G(N,L)$ and $G(N,p)$ ensembles, respectively. The Gibbs entropy in the $G(N,L)$ ensemble is given by [8]

$$N\Sigma = \log\begin{pmatrix} \frac{N(N-1)}{2} \\ L \end{pmatrix}. \tag{8}$$

As mentioned earlier, the corresponding probability of each link in the conjugate $G(N,p)$ ensemble is given by $p_{ij}=p=2L/[N(N-1)]$. Inserting this probability in the definition of the Shannon entropy [Eq. (7)] we get

$$\Sigma = S/N + \frac{1}{2N}\left[\log\left(\frac{N(N-1)}{2L}\right) - \log\left(\frac{N(N-1)}{2} - L\right)\right].$$

Therefore the Gibbs entropy Σ and the Shannon entropy per node S/N of random graphs are equal in the thermodynamic limit $N\to\infty$.

V. COMPARISONS BETWEEN THE NETWORK ENSEMBLES WITH GIVEN DEGREE SEQUENCE AND STRUCTURAL CUTOFF

The microcanonical ensemble of networks with given degree sequence $\{\kappa_i\}$ has been fully characterized in [9]. For simplicity, we consider the Gibbs entropy per node Σ in the case where the maximal connectivity of the nodes satisfy a structural cutoff, i.e., $k_{\max} < \sqrt{\langle\kappa\rangle N}$. In this limit the statistical mechanics treatment gives the Bender formula [18] and the Gibbs entropy per node Σ is given by

$$N\Sigma = \log[(2L-1)!!] - \sum_i \log(\kappa_i!) - \frac{1}{4}\left(\frac{\sum_i \kappa_i^2}{\sum_i \kappa_i}\right)^2. \tag{9}$$

In the conjugate-canonical ensemble, the probability of having a link is given by $p_{ij}=\kappa_i\kappa_j/\langle\kappa\rangle N$. Inserting this expression into Eq. (7) we get for the Shannon entropy of the ensemble

$$\Sigma = S/N - \sum_i \{\log[\kappa_i!/(\kappa_i^{\kappa_i}e^{-\kappa_i})]\} + O(\log(N)/N). \tag{10}$$

We observe that the entropy per node Σ and the Shannon entropy per node S/N of the canonical conjugated network ensemble are not equal in the thermodynamic limit. This implies, for example, that the entropy per node of regular networks is smaller than that of a Poisson network with same average degree. In particular, suppose we take, for regular networks, $\kappa_i=c$ and for the conjugated-canonical Poisson network, k_i to be a Poisson distributed random variable with a mean $\langle k_i\rangle=\kappa_i=c$. The entropy of regular networks Σ_R and the entropy of Poisson networks S_{ER} are related by the expression

$$\Sigma_R = S_{ER}/N - \log\frac{c!}{c^c e^{-c}} \simeq S_{ER}/N - \frac{1}{2}\log(c), \tag{11}$$

where in the last expression we have taken the Stirling approximation valid for large c.

The nonequivalence of Σ and S/N in the thermodynamic limit can be also checked for network ensembles satisfying further constraints as, for example, the networks ensembles with given degree sequence and network community structure and network ensembles with given degree sequence and given spatial dependence of the networks on the distance between the nodes. In general it is possible to demonstrate that as soon as we consider ensembles of networks with an extensive number of constraints the Gibbs entropy per node Σ and the Shannon entropy per node S/N are nonequal in the thermodynamic limit.

VI. VON NEUMANN ENTROPY OF A NETWORK ENSEMBLE

In [12] the authors have shown that is possible to define a von Neumann entropy of a network. This entropy is constructed from a density matrix ρ associated with the network. The density matrix must be a positive semidefinite matrix with unitary trace. In order to construct a density matrix from a network, in [12] it is proposed to consider the matrix $\rho=L/\Sigma_{ij}\, a_{ij}$, where L is the Laplacian matrix of the network, with $L_{ij}=\Sigma_r\, a_{ir}\delta_{i,j}-a_{ij}$. The spectrum [19] of the Laplacian matrix is important for the stability of $O(n)$ models, synchronization properties of networks, and determining the scaling of the return times of random walk on the network [20]. Given ρ as specified above, we can calculate the average von Neumann entropy of an ensemble as

$$S_{VN} = -\langle \mathrm{Tr}\,\rho\log(\rho)\rangle_{\Pi}. \tag{12}$$

The von Neumann entropy is therefore related to the spectra of the Laplacian. The theoretical evaluation of the self-averaging spectra of the Laplacian of complex networks ensemble is a very challenging topic that has attracted recent interest in the statistical mechanics community [19]. Here we numerically explore how the von Neumann entropy S_{VN} is related to the Shannon entropy of canonical ensembles.

For $G(N,p)$ networks the average von Neumann entropy, S_{VN}, is an increasing function of the average connectivity, pN, while the Shannon entropy per node, S/N, has the typical bell-shape form given by Eq. (8), in the limit of large N. Therefore, for the $G(N,p)$ random graphs ensembles, the relation between S_{VN} and S is nonmonotonic when we vary the average connectivity $p(N-1)$. It is instructive to study the

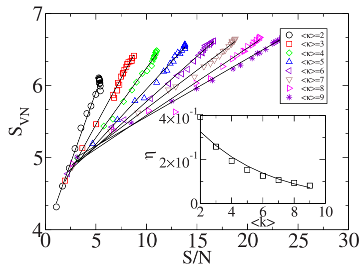

FIG. 1. (Color online) The von Neumann entropy S_{VN} versus the Shannon entropy per node S/N calculated for ensembles of scale-free networks with different expected average degree $\langle\kappa\rangle$. The points are calculated by averaging over 20 networks in the ensemble of networks with $N=1000$ nodes and different power-law exponents γ of the distribution of the expected degrees $P(\kappa)\propto\kappa^{-\gamma}$. The inset report the slope η defined in [13] as a function of $\langle\kappa\rangle$ and the exponential fit indicated as a solid line.

relation of the Shannon entropy of a network ensemble and its average von Neumann entropy in networks with the same average degree. In [9] it has been shown that networks with power-law degree distribution $P(k)\propto k^{-\gamma}$ and constant average degree $\langle k\rangle$ have a Gibbs entropy per node Σ which is an increasing function of the power-law exponent γ. Similarly the Shannon entropy per node S/N of canonical network ensembles with fixed expected degree κ_i, where $P(\kappa)\propto\kappa^{-\gamma}$ and fixed $\langle\kappa\rangle$ is increasing with the power-law exponent γ. Therefore changing the power-law exponent γ is a way to modulate S by leaving the average degree constant. In Fig. 1 we report the von Neumann entropy S_{VN} vs the Shannon entropy per node S/N in canonical power-law network ensembles with constant $\langle\kappa\rangle$ and variable value of the γ exponent. We find that the two entropies are linearly related

$$S_{VN}=\eta S/N+\beta, \tag{13}$$

where η decays exponentially as a function of $\langle\kappa\rangle$ for small values of $\langle\kappa\rangle\ll N$. Therefore for scale-free networks the von Neumann entropy is linearly related to the Shannon entropy of the canonical ensembles measuring the number of typical networks in the ensemble.

VII. CONCLUSIONS

In this Rapid Communication we have explored the connection between different definition of entropy of network ensembles. Interesting we have found that the Gibbs entropy per node Σ is equal to the Shannon entropy per node S/N in the thermodynamic limit for random graphs. However, when we consider networks with and extensive number of constraints (as, for example, a given degree distribution) the Gibbs entropy per node Σ and the Shannon entropy per node S/N differ by $\mathcal{O}(1)$ terms. Moreover we have related the Shannon entropy with the recently introduced von Neumann entropy of networks. Interestingly we found that for scale-free networks with constant average degree S_{VN} and S/N are linearly related. We believe that all the entropies of the network ensembles, S and S_{VN} as well as Σ [6] will play a crucial role for the quantification of the complexity and in inference problems in networks.

ACKNOWLEDGMENT

G.B. acknowledges stimulating discussions with J. Baranyi, A. C. C. Coolen, and A. N. Samukhin.

[1] R. Albert and A.-L. 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 the WWW* (Oxford University Press, Oxford, 2003); M. E. J. Newman, SIAM Rev. **45**, 167 (2003).

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

[3] S. N. Dorogovtsev, A. Goltsev, and J. F. F. Mendes, Rev. Mod. Phys. **80**, 1275 (2008); A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamics Processes on Complex Networks* (Cambridge University Press, Cambridge, 2008).

[4] S. Fortunato, e-print arXiv:0906.0612.

[5] D. Gfeller, J.-C. Chappelier, and P. De Los Rios, Phys. Rev. E **72**, 056135 (2005).

[6] G. Bianconi, P. Pin, and M. Marsili, Proc. Natl. Acad. Sci. U.S.A. **106**, 11433 (2009).

[7] T. Cover and J. Thomas, *Elements of Information Theory* (Wiley-Interscience, New York, 1991).

[8] L. Bogacz, Z. Burda, and B. Waclaw, Physica A **366**, 587 (2006).

[9] G. Bianconi, EPL **81**, 28005 (2008); Phys. Rev. E **79**, 036114 (2009).

[10] G. Bianconi, A. C. C. Coolen, and C. J. Perez Vicente, Phys. Rev. E **78**, 016114 (2008).

[11] J. Gomez-Gardenes and V. Latora, Phys. Rev. E **78**, 065102(R) (2008).

[12] S. L. Braunstein, S. Gosh, and S. Severini, Ann. Comb. **10**, 3 (2006); F. Passerini and S. Severini, e-print arXiv:0812.2597.

[13] C. Rovelli and F. Vidotto, e-print arXiv:0905.2983.

[14] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Muñoz, Phys. Rev. Lett. **89**, 258702 (2002); J. Park and M. E. J. Newman, Phys. Rev. E **70**, 066146 (2004).

[15] M. Boguñá and R. Pastor-Satorras, Phys. Rev. E **68**, 036112 (2003).

[16] D. Garlaschelli and M. I. Loffredo, Phys. Rev. E **78**, 015101(R) (2008).

[17] D. Garlaschelli and M. I. Loffredo, Phys. Rev. Lett. **102**, 038701 (2009).

[18] E. Bender and E. Rodney Canfield, J. Comb. Theory, Ser. A **24**, 296 (1978); B. D. McKay, Ars Comb. **19A**, 15 (1985).

[19] A. N. Samukhin, S. N. Dorogovtsev, and J. F. F. Mendes, Phys. Rev. E **77**, 036115 (2008); R. Kühn, J. Phys. A: Math. Theor. **41**, 295002 (2008); G. Bianconi, e-print arXiv:0804.1744.

[20] R. Burioni, D. Cassi, and A. Vezzani, Phys. Rev. E **60**, 1500 (1999); M. Barahona and L. M. Pecora, Phys. Rev. Lett. **89**, 054101 (2002).