Classical information theory of networks

To cite this article: Filippo Radicchi *et al* 2020 *J. Phys. Complex.* **1** 025001

View the article online for updates and enhancements.

Journal of Physics: Complexity

PAPER

Classical information theory of networks

Filippo Radicchi[1], Dmitri Krioukov[2], Harrison Hartle[2] and Ginestra Bianconi[3,4,5]

[1]   Center for Complex Networks and Systems Research, Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN 47408, United States of America

[2]   Department of Physics, Department of Mathematics, Department of Electrical & Computer Engineering, Northeastern University, Boston, MA 02115, United States of America

[3]   School of Mathematical Sciences, Queen Mary University of London, London, E1 4NS, United Kingdom

[4]   Alan Turing Institute, The British Library, London, United Kingdom

[5]   Author to whom any correspondence should be addressed.

E-mail: ginestra.bianconi@gmail.com

Abstract

Existing information-theoretic frameworks based on maximum entropy network ensembles are not able to explain the emergence of heterogeneity in complex networks. Here, we fill this gap of knowledge by developing a classical framework for networks based on finding an optimal trade-off between the information content of a compressed representation of the ensemble and the information content of the actual network ensemble. We introduce a novel classical network ensemble satisfying a set of soft constraints and we find the optimal distribution of the constraints for this ensemble. We show that for the classical network ensemble in which the only constraints are the expected degrees a power-law degree distribution is optimal. Also, we study spatially embedded networks finding that the interactions between nodes naturally lead to non-uniform spread of nodes in the embedding space, leading in some cases to a fractal distribution of nodes. This result is consistent with the so called 'blessing of non-uniformity' of data, i.e. the fact that real world data typically do not obey uniform distributions. The pertinent features of real-world air transportation networks are well described by the proposed framework.

1. Introduction

The principle of maximum entropy states that the unique probability distribution, encoding all the information available about a system but not any other information, is the one with largest information entropy [1]. Available information about the system corresponds to constraints under which entropy is maximized. The principle of maximum entropy has found applications in many different disciplines, including physics [2], computer science [3], geography [4], finance [5], molecular biology [6], neuroscience [7], learning [8], deep learning [9], etc.

Powerful information-theoretical frameworks that extend and generalize maximum entropy principles by making use of operations such as compression or erasure of information have been recently proposed. A paradigmatic example is the information bottleneck principle [8]. The principle allows to optimally learning a given output from an input signal, and the optimization relies on finding the best trade-off between accuracy of the prediction and effectiveness of the compression. Another notable example of this type of theoretical frameworks is the study of computation and the investigation of the entropic cost of bit erasure [10].

Applications of the maximum principle can be found also in network science [11–19], where the maximum entropy argument is applied to the distribution of probabilities $P(G)$ of observing a given graph G of finite size N in an ensemble of random graphs. Different entropy-maximization constraints lead to different network models. For example, if the constraints are soft, i.e., if they deal with expected values of network properties, then $P(G)$ is a Gibbs-like distribution corresponding to ERGMs [12, 20].

This approach can be used to model networks with heterogeneities, e.g., in node degrees [21, 22], edge weights [23], and community sizes [24, 25]. However, an important shortcoming of this approach is that it

cannot explain why these heterogeneities can be found so ubiquitously in real networks. Indeed, current maximum entropy approaches can only generate the least biased network ensembles with given expected degree sequence, but they cannot be used to explain or justify why in many cases we observe heterogeneous degree sequences. Similarly for spatially embedded networks, current maximum entropy approaches can be used to provide ensembles of spatial networks for a given distribution of nodes in the space, but they cannot be used to draw any conclusion on the expected spatial distribution of the nodes in the network. Therefore if we want to infer the positions of the nodes in network embedding we do not have information theory guidelines on how to choose the prior on the spatial distribution of the nodes.

In the present paper, we address this fundamental shortcoming of current information-theoretical approaches to the study of networks. Specifically, we derive a novel framework that is able to predict the optimal degree distribution and the optimal spatial distribution of nodes in space. Both distributions turn out to be heterogeneous, thus providing a principled explanation of the origin of heterogeneities in complex networks. Our approach is based on finding the best compressed representation of a network ensemble, given the content of information conveyed by the ensemble. We consider network ensembles where any pair of nodes is associated with a set of hidden variables obeying an arbitrary distribution, e.g., arbitrary degree distribution or arbitrary distribution of distances between pair of nodes, expressed in general as $P_\mathcal{V}(\mathbf{x})$. We measure the information content of the network ensemble and of its compressed network ensemble representation in terms the corresponding entropies S and H, respectively. Finally we propose to find the optimal hidden variable distribution $P_\mathcal{V}^\star(\mathbf{x})$, e.g., degree distribution or spatial distribution of distance between pair of nodes, by maximizing

$$P_\mathcal{V}^\star(\mathbf{x}) = \arg \max_{P_\mathcal{V}(\mathbf{x})}[H - \lambda S] \tag{1}$$

under the constraints that the network contains a given number of nodes and links, and that the entropy of the network ensemble is given, i.e., $S = S^\star$. As explained in the main text and in the appendices, this principle is solidly rooted in information theory [26] as the classical network ensemble and its compressed representation can be seen respectively as the input and output of a communication channel. Therefore the definition of the optimal hidden variable distribution can be interpreted as the optimal input distribution of a communication channel in information theory.

We believe that our results not only provide an information-theoretical explanation for the emergence of heterogeneous properties in complex networks, but also open a promising perspective for devising a new generation of inference methods for finding optimal network embeddings.

2. Results

2.1. Classical network ensembles

The simplest examples of maximum entropy ensemble are the $\mathcal{G}(N, p)$ and $\mathcal{G}(N, L)$ ensembles obtained by enforcing a constrain on the expected and the actual total number of links, respectively [27, 28]. In network theory these ensembles can be respectively generalized to canonical and microcanonical network ensembles enforcing a set of soft and hard constraints [13–15] which are not in general equivalent [14, 15, 29]. A major example of canonical network ensemble is the exponential random graph mode (ERGM) enforcing a given expected degree sequence [12] whose conjugated microcanonical ensemble is the configuration model [14, 15, 30].

In all the examples above, the maximum entropy principle is *de facto* applied to network adjacency matrices A whose elements are understood as sets of edge variables correlated by the imposed constraints. Calculations generally lead to the derivation of the probability $\pi_{ij} = P(A_{ij} = 1)$ for the pair of nodes i and j to be connected. If networks are undirected, then $A_{ij} = A_{ji}$ and $\pi_{ij} = \pi_{ji}$. This approach is very similar to the one used in quantum statistical mechanics to describe systems of noninteracting particles whose role is played by network edges, while particle states are enumerated by node pairs (i, j) [12, 13]. In fact the adjacency matrix element A_{ij} indicating the number of links between a pair of nodes (i, j) corresponds to the 'occupation number' in quantum statistical mechanics. Indeed in binary networks, where A_{ij} is either 0 or 1, π_{ij} takes the Fermi–Dirac form; if multiple edges are allowed between the same pair of nodes, then the system is described by the Bose–Einstein statistics [31].

Here we take advantage of the principle of maximum entropy in a classical way. Instead of dealing with all elements of the adjacency matrix (corresponding to the occupation numbers of quantum statistical mechanics) we look directly at network edges (which corresponds to particle states). Therefore a given network G of N nodes is identified by its edge list $\left\{\vec{\ell}^{[n]}\right\}$ with $n \in \{1, 2, \ldots, L\}$ where each link $\vec{\ell}^{[n]}$ is described by an ordered pair of node labels $\vec{\ell}^{[n]} = \left(\ell_1^{[n]}, \ell_2^{[n]}\right)$, i.e., $\ell_1^{[n]}$ indicates the label of the node $i \in \{1, 2, \ldots, N\}$ attached to the

first end of the link n and similarly $\ell_2^{[n]}$ indicates the label of the node attached to the second end of the link n (see appendices for details).

We assume that the ends of each link are drawn independently from the probability distribution $P(\vec{\ell})$. Therefore $P(\vec{\ell})$ indicates the probability that, by picking a random edge, nodes ℓ_1 and ℓ_2 are found at its ends. The Shannon entropy S of this ensemble is given by

$$S = -L\sum_{\vec{\ell}} P(\vec{\ell}) \ln\ P(\vec{\ell}). \tag{2}$$

S is named the classical entropy and quantifies the information content associated with all edges in the network. If we indicate with $\langle k \rangle$ average degree of the network, equation (2) indicates that the entropy S is given by the sum of $L = \langle k \rangle N/2$ identical terms corresponding to the entropy associated with the typical number of ways in which we can choose two nodes (i, j) to be connected by a single link.

The distribution $P(\vec{\ell})$ that describes the ensemble is then found using the maximum entropy principle. Different constraints in the entropy maximization problem lead to different distributions $P(\vec{\ell})$. Since the marginal probabilities in this ensemble are exponential, we refer to it as the classical network ensemble, differentiating it from previously explored maximum entropy ensembles where the marginals obey quantum statistics [12]. We note that the framework we consider here allows for multiedges and tadpoles as in similar approaches [4, 19]. This makes all edges uncorrelated variables, allowing for greater simplicity and flexibility.

2.2. Classical network ensemble with expected degrees

As the first very basic example of classical network ensemble, we consider the ensemble in which we constrain expected values of node degrees. That is, we require that the probability to find node i at one of the ends of a randomly chosen link is k_i/L,

$$\sum_{\vec{\ell}} P(\vec{\ell})\ [\mathbb{1}(\ell_1 = i) + \mathbb{1}(\ell_2 = i)] = \frac{k_i}{L}, \tag{3}$$

where $\{k_i\}$ is any given degree sequence, L is a fixed number of links in the network, which is assumed to be consistent with k_is via $2L = \sum_i k_i$, and where $\mathbb{1}(x = y)$ is the indicator function: $\mathbb{1}(x = y) = 1$ if $x = y$ and $\mathbb{1}(x = y) = 0$ otherwise. The constraint in equation (3) is required to hold for all nodes $i = 1, \ldots, N$.

The maximum entropy distribution $P(\vec{\ell})$ is found by maximizing the functional

$$\mathcal{G} = S - \mu L \left[\sum_{\vec{\ell}} P(\vec{\ell}) - 1\right] - L\sum_{i=1}^{N} \psi_i \left[\sum_{\vec{\ell}} P(\vec{\ell})\ [\mathbb{1}(\ell_1 = i) + \mathbb{1}(\ell_2 = i)] - \frac{k_i}{L}\right]. \tag{4}$$

where we have introduced the Lagrange multipliers ψ_i and μ associated with the constraint in equation (3) and the normalization of $P(\vec{\ell})$, respectively. The solution of this maximization problem leads to the expression for the probability π_{ij} that a given link connects node i at one end to node j at the other end, that is

$$\pi_{ij} = P(\ell_1 = i, \ell_2 = j) = e^{-\mu}e^{-\psi_i - \psi_j}, \tag{5}$$

where the Lagrange multipliers ψ_i and μ are the solutions of the constraint equations

$$\frac{k_i}{L} = 2\frac{k_i}{\langle k \rangle N} = 2e^{-\mu}e^{-\psi_i}\sum_{j=1}^{N} e^{-\psi_j}. \tag{6}$$

Therefore, $e^{-\psi_i} = k_i$ and $e^{\mu} = (\langle k \rangle N)^2$, from which we obtain

$$\pi_{ij} = \frac{k_i\,k_j}{(\langle k \rangle N)^2}. \tag{7}$$

Notice that π_{ij} is the probability that a link connects node i at the first end and node j at the second end, therefore the π_{ij} is a distribution and obeys the normalization condition $\sum_{ij}\pi_{ij} = 1$. Since there are $L = \langle k \rangle N/2$ links in the network, and two nodes are connected if there is a link attached to the two ends in any possible order, the average number of links that connect node i to node j is given by

$$\langle A_{ij} \rangle = 2L\pi_{ij} = \frac{k_i k_j}{\langle k \rangle N}. \tag{8}$$

This is the average number of links between nodes of degrees k_i and k_j in uncorrelated random networks [30]. Equation (8) is the starting point of many calculations in network science that use the uncorrelated random

networks as a null model. A popular example is the modularity function used in community detection [32]. The derivation above provides a theoretical ground for such an interpretation of the model.

We now turn to more sophisticated outcomes of the considered framework. In our classical network ensemble, the degree distribution $P(k)$ is an input parameter that we can set to whatever we wish. Among all possible choices of the degree distribution, which one corresponds to maximal randomness?

To answer this question, we note that we can express the classical network entropy S in terms of the degree distribution $P(k)$ as

$$S = -L\sum_{ij}\pi_{ij} \ln\ \pi_{ij} = \langle k\rangle N \left[\ln(\langle k\rangle N) - \sum_k \frac{kP(k)}{\langle k\rangle}\ln\ k\right]. \tag{9}$$

The entropy S quantifies the amount of information encoded in the classical network ensemble with N nodes, L edges, and degree distribution $P(k)$. Any given $P(k)$ uniquely determines the value of S via equation (9), yet the same value of S may correspond to different $P(k)$s.

2.3. Information theory framework

Here we describe our theoretical framework to predict the optimal degree distribution in terms of a standard information-theoretic problem [26]. A network instance is a 'message'. Specifically, a message consists of L two-letter words, each representing a link $\vec{\ell} = (i,j)$. Letters are node labels, so that the alphabet is given by N distinct symbols. We assume that messages are generated by picking random pairs of nodes according to the probability π_{ij}. For the classical network ensemble enforcing expected degrees the probability π_{ij} is only dependent on the degrees of the nodes, i.e.,

$$\pi_{ij} = \pi_{k_i,k_j} \tag{10}$$

with

$$\pi_{k,k'} = \frac{kk'}{(\langle k\rangle N)^2}, \tag{11}$$

indicating the probability that two nodes of degree k and k' are connected to one or the other end node of a link in the classical network ensemble. This is our source of messages. If we change degree sequence, then we have a different source of messages. The entropy S defined in equation (9) is the entropy of the source. In our specific setting, S turns out to be a function of the degree distribution $P(k)$ only, not of the specific degree sequence $\{k_1,\ldots,k_N\}$. Thus, if we change the degree distribution $P(k)$, then we change the source of messages.

Once generated, messages are compressed using a lossy compression channel. The choice of the channel is naturally suggested by the classical network ensemble under consideration. Since for the classical network ensemble enforcing expected degrees the probability $\pi_{ij} = \pi_{k_i,k_j}$ only depends on the two node degrees, we use the channel where the link labels (i,j) are replaced with the link label of the pair of degrees (k,k') of the two linked nodes. Please note that the messages are still the same as those generated by the source. However, many of them are no longer distinguishable after the application of the channel. Specifically the channel is erasing information about the actual identity of the linked nodes and is retaining only the information about their degrees.
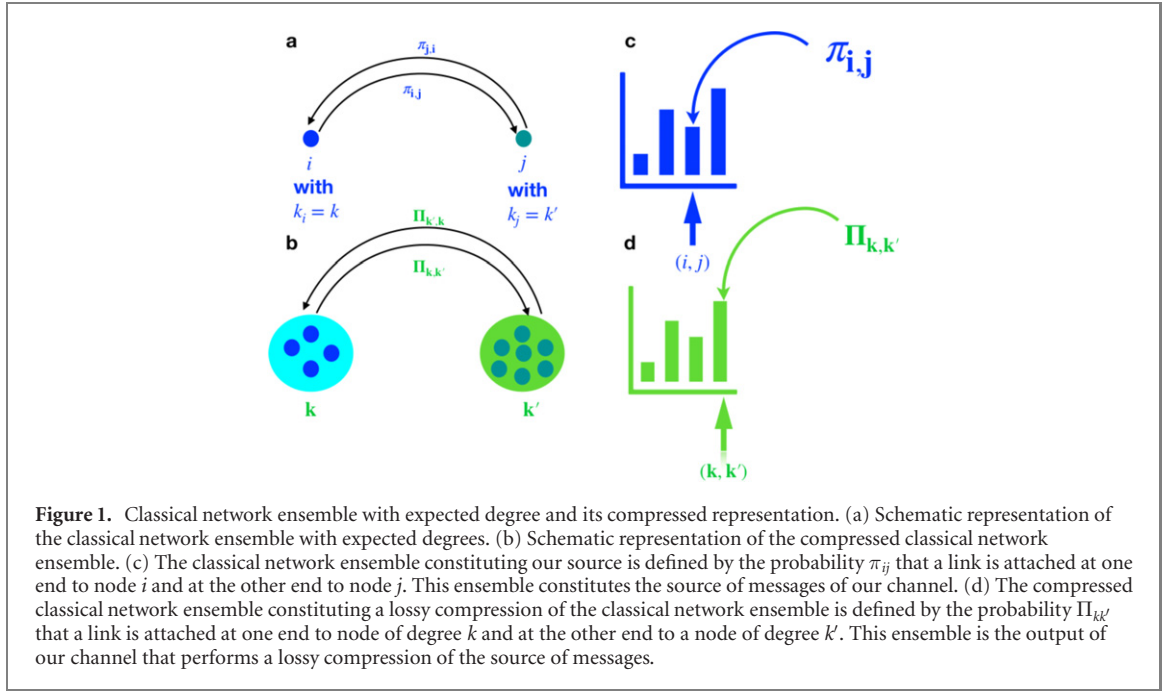
The output of the channel corresponds to a coarse-grained network ensemble (see figure 1), where all nodes with the same degree class are indistinguishable and they form a super node in the coarse-grained description. The network ensemble can be used to compress the information of the original network retaining only the information regarding the degree of the linked nodes. Clearly, in this ensemble we observe the same expected number of links $L_{k,k'}$ between nodes of degree k and nodes of degree k' as in the classical network ensemble. If we indicate with N_k the number of nodes in degree class k, it is easy to show that $L_{k,k'} = L\pi_{k,k'}N_kN_{k'}$ and that $\sum_{k,k'} L_{k,k'} = L$. Every link of the coarse-grained ensemble has probability $\Pi_{k,k'}$ to connect super-nodes corresponding to degree classes k and k' where

$$\Pi_{k,k'} = \frac{L_{k,k'}}{L} = \frac{kk'P(k)P(k')}{\langle k\rangle^2}. \tag{12}$$

This compressed ensemble is on its own a classical network ensemble, therefore its entropy H is

$$H = -L\sum_{k,k'}\Pi_{k,k'} \ln\ \Pi_{k,k'} = -\langle k\rangle N\sum_k \frac{kP(k)}{\langle k\rangle}\ln\left(\frac{kP(k)}{\langle k\rangle}\right). \tag{13}$$

We have two representations of the network ensemble at the node level and at the compressed level whose information content is quantified respectively by the S-entropy and the H-entropy. Note that the different notation is only introduced to distinguish between the entropy of the original ensemble and the entropy of its compressed version. However, the entropy H is nothing else that the entropy of a classical network ensemble

Figure 1. Classical network ensemble with expected degree and its compressed representation. (a) Schematic representation of the classical network ensemble with expected degrees. (b) Schematic representation of the compressed classical network ensemble. (c) The classical network ensemble constituting our source is defined by the probability π_{ij} that a link is attached at one end to node i and at the other end to node j. This ensemble constitutes the source of messages of our channel. (d) The compressed classical network ensemble constituting a lossy compression of the classical network ensemble is defined by the probability $\Pi_{kk'}$ that a link is attached at one end to node of degree k and at the other end to a node of degree k'. This ensemble is the output of our channel that performs a lossy compression of the source of messages.

whose nodes are degree classes. Given that our channel is only erasing information, we have the interesting results that the entropy H of the output of the channel is equal to the mutual information between the input and the output of the channel (see appendices) and represents a metric of effectiveness of the channel: the higher its value, the more effective is the channel in transmitting the information produced by the source.

In summary, we have potentially many sources of messages given by classical network ensemble with different $P(k)$s, but we have one given channel prescribed by our coarse-grained procedure of the network.

The maximization problem that we solve consists in determining the best distribution of hidden variables that maximizes the capacity of our channel for given value of the entropy S of the source. Therefore we maximize H for fixed value of S. The constraint on the entropy S is imposed as we do not want to compare the performance of the channel over arbitrary sources of messages, but only over sources with similar level of information.

The optimal degree distribution that will allow the best reconstruction of the original network ensemble given only the knowledge of its compressed representation is given by

$$P^\star(k) = \arg\max_{P(k)}[H - \lambda S], \tag{14}$$

where the optimization is performed under the constraints the network contains a given number of nodes and links and that $S = S^\star$.

We stress that our problem is formulated as essentially an optimization of the capacity of the channel aimed at finding the optimal distribution of hidden variables for any fixed value of the entropy $S = S^\star$ (see appendix A for more details about the oretical framework).

2.4. Optimal degree distribution

We now show how our theoretical framework can allow us to predict the optimal degree distribution of the classical network ensemble with expected degrees. We impose the constraint $S = S^\star$, where $S^\star$ is a given positive real number, i.e., we consider different network ensembles that have the same information content or 'explicative power' at the node level. To find the typical degree distribution $P(k)$ under this constraint, we maximize the randomness of the coarse-grained model quantified by the H-entropy. Clearly, $P(k)$ must also satisfy the constraints $\sum_k kP(k) = \langle k \rangle$ and $\sum_k P(k) = 1$. Combining all together, we thus have to maximize the functional

$$\mathcal{F} = H - \lambda\left[S - S^\star\right] - \mu N\left[\sum kP(k) - \langle k \rangle\right] - \nu N\left[\sum_k P(k) - 1\right], \tag{15}$$

from which we obtain

$$P^\star(k) = \langle k \rangle e^{-(\mu+1)} e^{-\nu\frac{\langle k \rangle}{k}} k^{-(\lambda+1)}. \tag{16}$$

The Lagrange multipliers λ, μ, and ν are determined by the imposed constraints and they always exist as long as $\lambda > 1$.

Figure 2. Entropy H as a function of $S^\star$. (a) The H-entropy in equation (13) is evaluated for the degree distribution $P(k)$ that maximizes the functional $\mathcal{F}$ in equation (15). We consider $N = 10^4$, and different values of the S-entropy constraint $S^\star$ and the average degree $\langle k \rangle$. (b) Same as in panel (a), but for the spatial ensemble H (21) and $\mathcal{F}$ (22) with the power-law linking probability $f(\delta) = \delta^{-\alpha}/z$. We consider $\alpha = 3$ and different values of z.

Figure 3. Schematic representation of the classical spatial ensembles and their compressed ensemble. (a) The classical spatial ensemble is defined by the probability π_{ij} that a link connects node i at one end and node j at the other end. (b) The compressed representation of the ensemble in panel A is defined by the probability $\Pi_{k,k',\delta}$ that a link connects a node of degree k at one end and a node of degree k' at distance δ from the first node at the other end.

Equation (16) shows that the optimal degree distribution $P^\star(k)$ with a given value of the classical entropy in equation (9) is a power law. To be precise, the power-law decay holds for large degrees k, while in the low-k region there is an exponential cutoff that affects the mean of the distribution. This result indicates the information theory benefits for the widespread presence of scale-free topologies in complex networks [21]. In figure 2(a) we show the entropy H as a function of $S^\star$ for different values of the average degree $\langle k \rangle$. The lower the $S^\star$, and consequently the lower the power-law exponent λ, the higher the entropy H. This is because even though the number of networks with a given degree sequence decreases as $S^\star$ and λ go down, the number of ways to split L links into classes of links connecting nodes of degrees k and k' increases. Therefore this result highlights the entropic benefit to have networks with broad (i.e., low λ values) degree distributions that correspond to low values of the S-entropy but to high values of the H-entropy. Interestingly, the same result could be obtained by maximizing the randomness of the classical network ensemble, and therefore optimizing the S-entropy while keeping fixed the informative power of its compressed description, i.e., the H-entropy.

2.5. Classical information theory of spatial networks

In the following, we apply the proposed information-theoretical approach to ensembles of spatial networks. We assume that networks are generated according to different 'sources' of messages rather than the classical network ensemble with expected degrees, and the lossy compression of the source of messages consists in replacing link labels (i, j) with link labels associated in the most general case to (k, k', δ) where k and k' are the degrees of the linked nodes and δ is their distance in the underlying space (see figure 3). The logic behind the formulation of the constrained maximization problem is still the same as above: we optimize sources corresponding to similar level of information for a specified and unchangeable channel.

Our goal is here to show how the proposed information theory approach can be used to predict the most likely distribution of the nodes in space when pairs of nodes have a given space-dependent linking probability. Our approach reveals that if nodes are interacting in a network, then interactions induce a natural tendency of the nodes to be distributed inhomogeneously in space. The finding is consistent with the so-called 'blessing of non-uniformity' of data, i.e., the fact that real-world data typically do not obey uniform distributions [33].

We first consider spatial networks without any degree constraints, and then combine spatial and degree-based information in heterogeneous spatial networks.

2.6. Space-dependent linking probability

Let δ_{ij} be the distance between nodes i and j in some embedding space, and $\omega(\delta)$ be the distance distribution between all the $\binom{N}{2}$ pairs of nodes, which we also call the correlation function: the number of pairs of nodes at distance δ is $\binom{N}{2}\omega(\delta) \simeq N^2\omega(\delta)/2$. We define a spatial classical network ensemble by imposing the constraint

$$\sum_{\vec{\ell}} P(\vec{\ell}) \left[\mathbb{1}(\ell_1 = i) + \mathbb{1}(\ell_2 = j) \right] F(\delta_{ij}) = c, \tag{17}$$

where $F(\delta)$ is a function of the distance. This constraint can be interpreted as a total 'cost' of the links. Different functions correspond to different ensembles. For example, in the ensemble with a cost of the link proportional to the their length, this function is $F(\delta) = \delta$. If it is $F(\delta) = \ln \delta$, then the cost of a links scales like the order of magnitude of link lengths. The maximum entropy principle dictates the maximization of the functional

$$\mathcal{G} = S - \mu[\sum_{\vec{\ell}} P(\vec{\ell}) - 1] - \alpha \left[\sum_{\vec{\ell}} P(\vec{\ell}) \left[\mathbb{1}(\ell_1 = i) + \mathbb{1}(\ell_2 = j) \right] F(\delta_{ij}) - c \right], \tag{18}$$

leading to

$$\pi_{ij} = \frac{f(\delta_{ij})}{N^2}, \tag{19}$$

with $f(\delta) = g(\delta)/z$, $z = \int d\delta\, \omega(\delta)g(\delta)$, and $g(\delta) = e^{-\alpha F(\delta)}$. Therefore if $F(\delta) = \ln \delta$, then the linking probability decays with the distance as a power law, $g(\delta) = \delta^{-\alpha}$. If $F(\delta) = \delta$, then this decay is exponential, $g(\delta) = e^{-\alpha\delta}$. Fixing the number of links in the network to L as before, the classical entropy of the ensemble is

$$S = -L\sum_{ij}\pi_{ij} \ln \pi_{ij} = \langle k \rangle N \ln N - \frac{1}{2}\langle k \rangle N \int d\delta\, \omega(\delta)f(\delta) \ln f(\delta), \tag{20}$$

which is the spatial analogue of the classical entropy in equation (9).

We now ask: what is the optimal distribution of nodes in the space at parity of explicative power of the network model? That is, what is the optimal correlation function $\omega^\star(\delta)$ for given value of the entropy $S = S^\star$? To answer this question, we define the entropy H of the compressed model in which we consider only the number of ways to distribute L links such that every link connects two nodes at distance δ with probability density Π_δ

$$H = -L \int d\delta\, \Pi_\delta \ln \Pi_\delta, \tag{21}$$

where $\Pi_\delta = L_\delta/L$ and $L_\delta = L\omega(\delta)f(\delta)$ is the expected number of links between nodes at distance δ in the classical network ensemble (which clearly satisfy the normalization condition $\int d\delta\, L_\delta = L$). Our information theory framework shows that the maximum entropy value of $\omega(\delta)$ is then found by maximizing the functional

$$\mathcal{F} = H - \lambda \left[S - S^\star \right] - \mu L \left[\int d\delta\, \omega(\delta)f(\delta) - 1 \right] - \nu L \left[\int d\delta\, \omega(\delta) - 1 \right], \tag{22}$$

where λ, μ and ν are the Lagrange multipliers coupled with the constraints. The solution reads

$$\omega^\star(\delta) = e^{-(\mu+1)}e^{-\nu/f(\delta)}f(\delta)^{-(\lambda+1)}, \tag{23}$$

so that L_δ is given by

$$L_\delta^\star = L\, e^{-(\mu+1)}e^{-\nu/f(\delta)}f(\delta)^{-\lambda}, \tag{24}$$

The Lagrange multipliers are then found as the solutions of the constraints equations. In figure 2(b), we show the entropy H as a function of $S^\star$ for a power-law decaying linking probability $f(\delta) = \delta^{-\alpha}/z$.

We now make several important observations. First, if the space has no boundary, and is isotropic and homogeneous, then the networks are homogeneous since any two points in the space are equivalent and the linking probability depends only on the distance between pairs of points. The degree distribution is thus the Poisson distribution with the mean equal to the average degree $\langle k \rangle = 2L/N$. Second, equation (23) says that the maximum entropy distribution $\omega(\delta)$ of distances δ between the nodes in the space is uniquely determined by the linking probability $f(\delta)$. Third, if this probability decays as a power law $f(\delta) = \delta^{-\alpha}/z$, then the framework describes the natural emergence of power-law pair correlation functions. Specifically, the solution in equation (23) decays as a power law at small distances δ, while at large distances the decay is exponential due

to the finiteness of the system. If the embedding space is Euclidean of dimension d, then points are scattered in the space according to a fractal distribution. Define the node pair density function by

$$\rho(\delta) = \frac{\omega^\star(\delta)}{\Omega_\delta}, \tag{25}$$

where Ω_δ is the volume element at distance δ from an arbitrary point. In the d-dimensional Euclidean space, Ω_δ is the volume of the $(d-1)$-dimensional spherical shell, scaling with δ as $\Omega_\delta \propto \delta^{d-1}$. Therefore for a power-law linking probability $f(\delta) = \delta^{-\alpha}/z$, we get

$$\rho(\delta) \propto \delta^\beta e^{-\nu z \delta^\alpha}, \tag{26}$$

where $\beta = (\lambda+1)\alpha - (d-1)$. Therefore, the embedding in d dimensions is possible only if $\beta < 0$. Finally, the distribution of nodes in the space is fractal, and therefore highly nonuniform, as the uniform distribution would correspond to $\rho(\delta) = $ const.

2.7. Constraining expected values of node degrees and link costs

As the last example, we consider the classical network ensemble of spatial heterogeneous networks combining the degree and spatial constraints of equations (3) and (17), respectively. The probability π_{ij} that a random link connects nodes i and j is given by

$$\pi_{ij} = \frac{\kappa_i \kappa_j}{(\langle k \rangle N)^2} f(\delta_{ij}), \tag{27}$$

where $f(\delta) = e^{-\alpha F(\delta)}/z$, with α the Lagrangian multiplier coupled with the constraint in equation (17), z the normalization constant enforcing $\sum_{i,j} \pi_{ij} = 1$, and κ_i the hidden variable of node i given by $\kappa_i = e^{-\psi_i} \langle k \rangle N$, with ψ_i the Lagrangian multiplier coupled with the constraint in equation (3). If there are no correlations between the positions of the nodes in the space and their degrees, then the probability π_{ij} can be written as

$$\pi_{ij} = \frac{k_i k_j}{(\langle k \rangle N)^2} f(\delta_{ij}), \tag{28}$$

meaning that $\kappa_i = k_i$, so that κ_i can be interpreted as the expected degree k_i of node i. Using the same approximation as in reference [34] for a power-law decaying function $f(\delta_{ij}) = \delta_{ij}^{-\alpha}/z$, we can write π_{ij} as $\pi_{ij} \propto e^{-r_{ij}}$, where $r_{ij} = \ln \kappa_i + \ln \kappa_j - \alpha \ln \delta_{ij}$ is approximately the hyperbolic distance between nodes i and j located at radial coordinates $\ln \kappa_i$ and $\ln \kappa_j$ and at the angular distance proportional to δ_{ij}. Parameter α can then be related to the hyperbolic space curvature. The classical entropy of this ensemble is given by

$$S = -L\sum_{ij} \pi_{ij} \ln \pi_{ij} = \langle k \rangle N \ln[\langle k \rangle N] - \frac{1}{2}\langle k \rangle N \int d\kappa \int d\kappa' \int d\delta \, \omega(\kappa, \kappa', \delta) \, \kappa\kappa' f(\delta) \ln[\kappa\kappa' f(\delta)] \tag{29}$$

where $\omega(\kappa, \kappa', \delta)$ is the density of pairs of nodes with hidden variables κ and κ' at distance δ.

What is the optimal pair correlation function $\omega^\star(\kappa, \kappa', \delta)$ for a fixed value of the classical entropy $S = S^\star$? To answer this question, we maximize the H-entropy of the compressed model

$$H = -L \int d\kappa \int d\kappa' \int d\delta \, \Pi_{\kappa,\kappa',\delta} \ln \Pi_{\kappa,\kappa',\delta}, \tag{30}$$
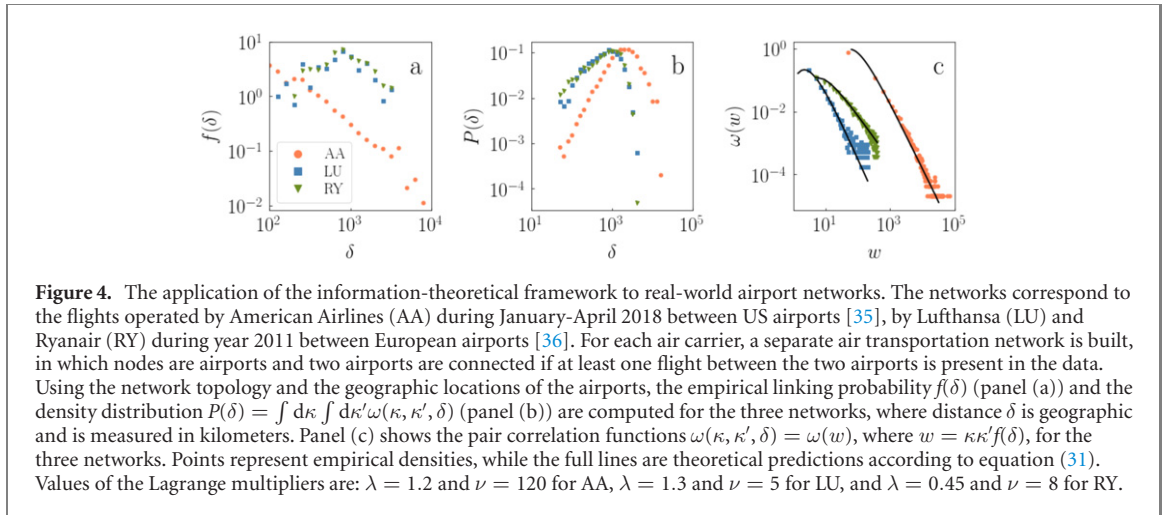
where $\Pi_{\kappa,\kappa',\delta} = L_{\kappa,\kappa',\delta}/L$ is the probability density that a link connected two nodes of with hidden variables κ and κ' and at distance δ. Note that $L_{\kappa,\kappa',\delta} = L\omega(\kappa, \kappa', \delta)\kappa\kappa' f(\delta)$ indicates the expected number of links between pairs of nodes with hidden variables κ and κ' at distance δ in the classical network ensemble. The maximization of H under the constraints $S = S^\star$, the normalization of $L_{\kappa,\kappa',\delta}$, $\int d\kappa \int d\kappa' \int d\delta \, L_{\kappa,\kappa',\delta} = L$, and the normalization of $\omega(\kappa, \kappa', \delta)$, $\int d\kappa \int d\kappa' \int d\delta \, \omega(\kappa, \kappa', \delta) = 1$, yields the answer

$$\omega^\star(\kappa, \kappa', \delta) = e^{-(\mu+1)} \exp\{-\nu/[\kappa\kappa' f(\delta)]\}[\kappa\kappa' f(\delta)]^{-(\lambda+1)}, \tag{31}$$

where λ, μ and ν are the Lagrange multipliers coupled with the $S = S^\star$ constraint, the normalization of $L_{\kappa,\kappa',\delta}$, and the normalization of $\omega(\kappa, \kappa', \delta)$, respectively. Observe that the pair correlation function $\omega(\kappa, \kappa', \delta)$ depends on its arguments only via $w = \kappa\kappa' f(\delta)$, and for small values of w it decays as a power-law function of w. If $f(\delta) = \delta^{-\alpha}/z$, then $\omega(\kappa, \kappa', \delta)$ can be also written in terms of the approximate hyperbolic distance $r = \ln w = \ln \kappa + \ln \kappa' - \alpha \ln \delta$ as

$$\omega^\star(r) = e^{-(\mu+1)} \exp[-(\lambda+1)r - \nu e^{-r}]. \tag{32}$$

As in the homogeneous spatial case, here we also observe that the optimal distribution of nodes in the space is not uniform.

Figure 4. The application of the information-theoretical framework to real-world airport networks. The networks correspond to the flights operated by American Airlines (AA) during January-April 2018 between US airports [35], by Lufthansa (LU) and Ryanair (RY) during year 2011 between European airports [36]. For each air carrier, a separate air transportation network is built, in which nodes are airports and two airports are connected if at least one flight between the two airports is present in the data. Using the network topology and the geographic locations of the airports, the empirical linking probability $f(\delta)$ (panel (a)) and the density distribution $P(\delta) = \int d\kappa \int d\kappa' \omega(\kappa, \kappa', \delta)$ (panel (b)) are computed for the three networks, where distance δ is geographic and is measured in kilometers. Panel (c) shows the pair correlation functions $\omega(\kappa, \kappa', \delta) = \omega(w)$, where $w = \kappa\kappa' f(\delta)$, for the three networks. Points represent empirical densities, while the full lines are theoretical predictions according to equation (31). Values of the Lagrange multipliers are: $\lambda = 1.2$ and $\nu = 120$ for AA, $\lambda = 1.3$ and $\nu = 5$ for LU, and $\lambda = 0.45$ and $\nu = 8$ for RY.

2.8. Real-world networks

In figure 4 we apply the considered information-theoretic framework to real-world air transportation networks, in which nodes are airports and edges between pairs of nodes indicate the existence of at a least one flight connecting the two airports. Specifically, we consider three networks corresponding to flights operated in different geographic areas by three air carriers. The distances δ_{ij} between airports i and j are their geographic distances. The linking probability $f(\delta)$ is computed from the data as the empirical connection probability, and the hidden variables κ_i are set to the actual degrees of the airports in the networks. We note that the empirical connection probabilities $f(\delta)$ decay as power laws, and that the pair correlation functions $\omega(w)$ are well described by equation (31).

3. Discussion

In summary, this work illustrates a classical information-theoretical approach to the characterization of random networks. This framework is based on a tradeoff between the entropy of the network ensemble and the entropy of its compressed representation. According to our theory, network inhomogeneities in the distribution of node degrees and/or node position in space both emerge from the general principle of maximizing randomness at parity of explicative power. The framework provides theoretical foundations for a series of models often encountered in network science, and can likely be extended to generalized network models such as multilayer networks and simplicial complexes [37, 38] or to information theory approaches based on the network spectrum [39]. In applications to real-world networks, the framework provides a theoretical explanation of the nontrivial inhomogeneities that are an ubiquitous features of real-world complex systems.

Data availability

The networks considered in this study are generated from data corresponding to flights operated by American Airlines (AA) during January-April 2018 between US airports [35], and by Lufthansa (LU) and Ryanair (RY) during year 2011 between European airports [36]. Geographical coordinates of the airports have been obtained from https://openflights.org/data.html. AA data are available upon request. LU and RY data can be downloaded from the repository Air Transportation Multiplex at http://complex.unizar.es/atnmultiplex/.

Code availability

Code used to generate figures 2 and 4 is available upon request.

Acknowledgments

Appendix A. The classical network ensemble

We consider a classical network ensemble defining the probability of a network $G = (V, E)$ of $|V| = N$ nodes and $|E| = L$ links. In this ensemble, a network G is described by an edge list $\left\{ \vec{\ell}^{\,[n]} \right\}$ with $n \in \{1, 2, \ldots, L\}$ where each link $\vec{\ell}^{\,[n]}$ is described by an ordered pair of node labels $\vec{\ell}^{\,[n]} = \left(\ell_1^{[n]}, \ell_2^{[n]} \right)$, i.e., $\ell_1^{[n]}$ indicates the label of the node $i \in \{1, 2, \ldots, N\}$ attached to the first end of the link n and similarly $\ell_2^{[n]}$ indicates the label of the node attached to the second end of the link n.

Every link variable $\ell^{[n]}$ can assume values of the type $(i^{[n]}, j^{[n]})$ with $i, j \in \{1, 2, \ldots, N\}$. In the classical network ensemble, every link is independently distributed, thus we associate to each network (edge list) $\{\vec{\ell}^{\,n}\}$ a probability

$$\mathbb{P}(G) = \prod_{n=1}^{L} P\left(\vec{\ell}^{\,[n]} \right) \tag{A.1}$$

where $P(\vec{\ell}^{\,[n]})$ is the probability that the nth link is connected to the pair of nodes $(\ell_1^{[n]}, \ell_2^{[n]})$. The entropy of this ensemble is given by

$$S(G) = -L \sum_{\vec{\ell}} P(\vec{\ell}) \ln P(\vec{\ell}). \tag{A.2}$$

Note that alternatively we could define the network ensemble as given by a set of L undistinguishable links defined as unordered pairs of node labels $\vec{\ell}$. In that case, by following similar mathematical steps as those used to treat the Gibbs paradox [40] in statistical mechanics, the entropy would only differ by a constant term, i.e.,

$$S^{[\text{undis}]}(G) = -\sum_{\{\vec{\ell}\}} \mathbb{P}(\{\vec{\ell}\}) \ln \mathbb{P}(\{\vec{\ell}\})$$

$$= -L \sum_{\vec{\ell}} P(\vec{\ell}) \ln P(\vec{\ell}) - \ln(L! 2^L). \tag{A.3}$$

The above entropy might be preferred to the entropy S associated to distinguishable links. However, the S and $S^{[\text{undis}]}$ entropies differ only by a global term that depends on the total number of links only, thus making S and $S^{[\text{undis}]}$ equivalent for the purpose of our mathematical framework. We further note that the classical network ensemble is fully described by the link ensemble. The link ensemble is a triple $(\vec{\ell}, \mathcal{A}_{\vec{\ell}}, \mathcal{P}_{\vec{\ell}})$ where $\vec{\ell}$ indicates the value associated of the random variable associated to an arbitrary link of the network, $\mathcal{A}_{\ell} = \{(i, j) | i, j \in \{1, 2, \ldots, N\}\}$ indicates the set of all distinct possible values that the link random variable can assume, and $\mathcal{P}_{\ell} = \{\pi_{ij}, i, j \in \{1, 2, \ldots, N\}\}$ indicates the set of probabilities

$$\mathbb{P}(\vec{\ell} = (i, j)) = \pi_{i,j}. \tag{A.4}$$

Here, we consider maximum entropy classical network ensembles where the probabilities π_{ij} only depend on some hidden variables $\mathbf{x_{ij}}$ associated to the link, i.e., where

$$\pi_{ij} = \pi_{\mathbf{x_{ij}}}. \tag{A.5}$$

Alternatively, we could say that $\pi_{i,j}$ is the probability that $\vec{\ell} = (i, j)$ given that the two nodes are characterized by the hidden variables $\mathbf{x}_{ij}$ assigned *a priori* to each pair of nodes of the network. For instance, if we consider the classical ensemble in which we constraint the expected degree sequence, we will have

$$\pi_{i,j} = \pi_{k_i, k_j}. \tag{A.6}$$

while in the spatial network we will have

$$\pi_{i,j} = \pi_{\kappa_i, \kappa_j, \delta_{ij}}. \tag{A.7}$$

Thus, the entropy $S(G)$ of the classical network ensemble is given by

$$S(G) = -L \sum_{i,j} \pi_{ij} \ln \pi_{ij} = -L \sum_{\mathbf{x}} N^2 P_{\mathcal{V}}(\mathbf{x}) \pi_{\mathbf{x}} \ln \pi_{\mathbf{x}}, \tag{A.8}$$

where

$$P_{\mathcal{V}}(\mathbf{x}) = \frac{1}{N^2} \sum_{i,j} \delta(\mathbf{x}_{ij}, \mathbf{x}) \tag{A.9}$$

indicates the probability that a random pair of nodes has hidden variable $\mathbf{x}_{ij} = \mathbf{x}$. Since the network ensemble is constructed given the distribution of hidden variables, the entropy $S(G)$ can be interpreted as a conditional

entropy of the network given the distribution of hidden variables as the rightmost term of equation (A.8) reveals.

Appendix B. The channel that compresses information

It follows that a classical network ensemble can be considered as a source of L messages. Each message is a link $\vec{\ell}$ carrying information on the node labels of the two linked nodes. We assume that the information is compressed by a channel Q, characterized by an input $\vec{\ell}$ taking values in $\mathcal{A}_{\vec{\ell}}$ and an output $\mathbf{x}(\vec{\ell})$ indicating the hidden variables associated to the link

$$\mathbf{x}(\vec{\ell}) = \mathbf{x}_{ij}. \tag{B.1}$$

The channel Q is a lossy compression channel that is erasing information about the identity of the nodes, and retaining only the value of their hidden variables. The output of the channel Q is the ensemble $\{\mathbf{x}(\vec{\ell}), \mathcal{A}_{\mathbf{x}}, \mathcal{P}_{\mathbf{x}}\}$, where the random variables associated to each link are given by the hidden variables of the linked nodes $\mathbf{x}(\ell)$. $\mathcal{A}_{\mathbf{x}}$ is the set of all possible values that the hidden variables of a link can take, and $\mathcal{P}_{\mathbf{x}}$ is the set of all probabilities

$$\Pi_{\bar{\mathbf{x}}} = \mathbb{P}\left(\mathbf{x}(\vec{\ell}) = \bar{\mathbf{x}}\right) = \sum_{i<j} \pi_{ij} \delta\left(\mathbf{x}(\vec{\ell}), \bar{\mathbf{x}}\right) = \pi_{\bar{\mathbf{x}}} N^2 P_{\mathcal{V}}(\bar{\mathbf{x}}). \tag{B.2}$$

For instance, if the hidden variables are exclusively the expected degrees of the nodes, we have

$$\Pi_{k,k'} = \mathbb{P}(\mathbf{x}(\vec{\ell}) = (k, k')) = \sum_{i<j} \pi_{ij} \delta(k_i, k) \delta(k_j, k') = \pi_{k,k'} N^2 P(k) P(k'). \tag{B.3}$$

If instead we are considering a spatial network with hidden variables $\mathbf{x} = (\kappa, \kappa', \delta)$, we have

$$\Pi_{k,k',\delta} = \mathbb{P}(\mathbf{x}(\vec{\ell}) = (k, k', \delta)) = \sum_{i<j} \pi_{ij} \delta(k_i, k) \delta(k_j, k') \delta(\delta_{ij}, \delta) = N^2 \pi_{k,k',\delta} \omega(\kappa, \kappa', \delta), \tag{B.4}$$

where here we have adopted the notation of the main text. The output message defines a compressed network ensemble of networks having entropy

$$H = -L \sum_{\mathbf{x}} \Pi_{\mathbf{x}} \ln \Pi_{\mathbf{x}} = -L \sum_{\mathbf{x}} N^2 \pi_{\mathbf{x}} P_{\mathcal{V}}(\mathbf{x}) \ln \left(N^2 \pi_{\mathbf{x}} P_{\mathcal{V}}(\mathbf{x})\right). \tag{B.5}$$

Interestingly, we have that the entropy H is equal to the mutual information between the input message and the output messages of the channel Q multiplied by L, i.e.,

$$H = L \sum_{\vec{\ell}, \mathbf{x}} P(\vec{\ell}, \mathbf{x}) \ln \left(\frac{P(\vec{\ell}, \mathbf{x})}{P(\vec{\ell}) P(\mathbf{x})}\right). \tag{B.6}$$

This fact follows immediately from the observation that the joint distribution $P(\vec{\ell}, \mathbf{x})$ between the source message $\vec{\ell} = (i, j)$ and the compressed message $\mathbf{x}$ is simply given by

$$P(\vec{\ell}, \mathbf{x}) = P(\vec{\ell}) \delta(\mathbf{x}_{ij}, \mathbf{x}), \tag{B.7}$$

i.e., the value of $\mathbf{x}(\vec{\ell})$ is uniquely determined by $\vec{\ell}$ and the relation

$$P(\mathbf{x}) = \sum_{\vec{\ell}} P(\vec{\ell}, \mathbf{x}) = \Pi_{\mathbf{x}}. \tag{B.8}$$

Appendix C. Optimal hidden variable distribution

Our framework aiming at finding the optimal distribution of hidden variables $P_{\mathcal{V}}^{\star}(\mathbf{x})$ consists in maximizing H for a fixed value of $S = S^{\star}$. In particular, we define the optimal hidden variable distribution $P_{\mathcal{V}}^{\star}(\mathbf{x})$ as the solution of the optimization problem

$$P_{\mathcal{V}}^{\star}(\mathbf{x}) = \arg\max_{P_{\mathcal{V}}(\mathbf{x})} [H - \lambda S]. \tag{C.1}$$

under the constraints the network contains a given number of nodes and links, and that $S = S^{\star}$. Since H is proportional to the mutual information of the channel Q, the maximum of H given $S = S^{\star}$ consists in the capacity of the channel under the constraint that the network ensemble has entropy $S = S^{\star}$. Interestingly, our

optimal hidden variable distribution can be seen as a parallel of the *optimal input distribution* [26] of a channel, with the difference that we consider a network model where $\pi_{\mathbf{x}}$ fixed and we optimize only the distribution of the hidden variables $P_{\mathcal{V}}(\mathbf{x})$.

ORCID iDs

Ginestra Bianconi https://orcid.org/0000-0002-3380-887X

References

[1] Jaynes E T 1957 Information theory and statistical mechanics *Phys. Rev.* **106** 620
[2] Pressé S, Ghosh K, Lee J and Dill K A 2013 Principles of maximum entropy and maximum caliber in statistical physics *Rev. Mod. Phys.* **85** 1115
[3] Ratnaparkhi A 1997 A simple introduction to maximum entropy models for natural language processing *University of Pennsylvania Institute for Research in Cognitive Science Technical Report No. IRCS-97-08* (Philadelphia, PA: University of Pennsylvania) p 81
[4] Wilson A G 1967 A statistical theory of spatial distribution models *Transp. Res.* **1** 253–69
[5] Kanbur R and Zhang X 2005 Fifty years of regional inequality in china: a journey through central planning, reform, and openness *Rev. Dev. Econ.* **9** 87–106
[6] Shi J, Zhao J, Li T and Chen L 2019 Detecting direct associations in a network by information theoretic approaches *Sci. China Math.* **62** 823–38
[7] Tang A, Jackson D, Hobbs J, Chen W, Smith J L, Patel H, Prieto A, Dumitru P, Grivich M I and Sher A 2008 A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro *J. Neurosci.* **28** 505–18
[8] Tishby N, Pereira F C and Bialek W 2000 The information bottleneck method (arXiv:physics/0004057)
[9] Song J, Marsili M and Jo J 2018 Resolution and relevance trade-offs in deep learning *J. Stat. Mech.* 123406
[10] Wolpert D H 2019 The stochastic thermodynamics of computation *J. Phys. A: Math. Theor.* **52** 193001
[11] Cimini G, Squartini T, Saracco F, Garlaschelli D, Gabrielli A and Guido C 2019 The statistical physics of real-world networks *Nat. Rev. Phys.* **1** 58
[12] Park J and Newman M E J 2004 Statistical mechanics of networks *Phys. Rev. E* **70** 066117
[13] Bianconi G 2009 Entropy of network ensembles *Phys. Rev. E* **79** 036114
[14] Anand K and Bianconi G 2009 Entropy measures for networks: toward an information theory of complex topologies *Phys. Rev. E* **80** 045102
[15] Anand K and Bianconi G 2010 Gibbs entropy of network ensembles by cavity methods *Phys. Rev. E* **82** 011116
[16] Bianconi G 2007 A statistical mechanics approach for scale-free networks and finite-scale networks *Chaos* **17** 026114
[17] Peixoto T P 2012 Entropy of stochastic blockmodel ensembles *Phys. Rev. E* **85** 056122
[18] Horvát S, Czabarka É and Toroczkai Z 2015 Reducing degeneracy in maximum entropy models of networks *Phys. Rev. Lett.* **114** 158701
[19] Sagarra O, Vicente C J P and Dïaz-Guilera A 2013 Statistical mechanics of multiedge networks *Phys. Rev. E* **88** 062806
[20] Holland P W and Leinhardt S 1981 An exponential family of probability distributions for directed graphs *J. Am. Stat. Assoc.* **76** 33–50
[21] Barabási A-L and Albert R 1999 Emergence of scaling in random networks *Science* **286** 509–12
[22] Voitalov I, van der Hoorn P, van der Hofstad R and Krioukov D 2019 Scale-free networks well done *Phys. Rev. Res.* **1** 033034
[23] Barrat A, Barthelemy M, Pastor-Satorras R and Vespignani A 2004 The architecture of complex weighted networks *Proc. Natl Acad. Sci.* **101** 3747–52
[24] Radicchi F, Castellano C, Cecconi F, Loreto V and Parisi D 2004 Defining and identifying communities in networks *Proc. Natl Acad. Sci.* **101** 2658–63
[25] Lancichinetti A, Fortunato S and Radicchi F 2008 Benchmark graphs for testing community detection algorithms *Phys. Rev. E* **78** 046110
[26] MacKay D J C 2003 *Information Theory, Inference and Learning Algorithms* (Cambridge: Cambridge University Press)
[27] Erdős P and Rényi A 1959 On random graphs i *Publ. Math. Debrecen* **6** 290–7
[28] Bollobás B and Béla B 2001 *Random Graphs* (*Number 73*) (Cambridge: Cambridge university press)
[29] Squartini T, de Mol J, Frank den Hollander and Garlaschelli D 2015 Breaking of ensemble equivalence in networks *Phys. Rev. Lett.* **115** 268701
[30] Molloy M and Reed B 1995 A critical point for random graphs with a given degree sequence *Random Struct. Algorithm* **6** 161–80
[31] Garlaschelli D and Loffredo M I 2009 Generalized bose-fermi statistics and structural correlations in weighted networks *Phys. Rev. Lett.* **102** 038701
[32] Newman M E J and Girvan M 2004 Finding and evaluating community structure in networks *Phys. Rev. E* **69** 026113
[33] Domingos P M 2012 A few useful things to know about machine learning *Commun. ACM* **55** 78–87
[34] Krioukov D, Papadopoulos F, Kitsak M, Amin V and Boguná M 2010 Hyperbolic geometry of complex networks *Phys. Rev. E* **82** 036106
[35] Bureau of Transport Statistics (Washington, DC: United States Department of Transportation) (https://www.bts.gov)
[36] Cardillo A, Gómez-Gardenes J, Zanin M, Romance M, Papo D, Pozo F D and Boccaletti S 2013 Emergence of network features from multiplexity *Sci. Rep.* **3** 1344
[37] Bianconi G 2013 Statistical mechanics of multiplex networks: Entropy and overlap *Phys. Rev. E* **87** 062806
[38] Owen T C and Bianconi G 2016 Generalized network structures: the configuration model and the canonical ensemble of simplicial complexes *Phys. Rev. E* **93** 062311
[39] Domenico M D and Jacob B 2016 Spectral entropies as information-theoretic tools for complex network comparison *Phys. Rev. X* **6** 041062
[40] Huang K 1963 *Statistical Mechanics* (New York: Wiley)