

Thermodynamics of network model fitting with spectral entropies

Carlo Nicolini,^{*} Vladimir Vlasov, and Angelo Bifone

Center for Neuroscience and Cognitive Systems, Istituto Italiano di Tecnologia, Corso Bettini 31, 38068 Rovereto (TN), Italy



(Received 29 January 2018; revised manuscript received 4 June 2018; published 31 August 2018)

An information-theoretic approach inspired by quantum statistical mechanics was recently proposed as a means to optimize network models and to assess their likelihood against synthetic and real-world networks. Importantly, this method does not rely on specific topological features or network descriptors but leverages entropy-based measures of network distance. Entertaining the analogy with thermodynamics, we provide a physical interpretation of model hyperparameters and propose analytical procedures for their estimate. These results enable the practical application of this novel and powerful framework to network model inference. We demonstrate this method in synthetic networks endowed with a modular structure and in real-world brain connectivity networks.

DOI: [10.1103/PhysRevE.98.022322](https://doi.org/10.1103/PhysRevE.98.022322)

I. INTRODUCTION

Many natural and artificial phenomena can be represented as networks of interacting elements. The mathematical framework of network theory can be applied across disciplines, ranging from sociology to neuroscience, and provides a powerful means to investigate a variety of diverse phenomena [1]. Unveiling the structural and organizational principles of complex networks often implies comparison with statistical network models. Generative models, for example, describe mechanisms of network wiring and evolution [2,3] or the constraints that may have contributed to shaping the network topology during its development [4]. Null models are used to describe maximally random networks with specific features, for example, prescribed sequences of node degrees [5].

Maximum-likelihood approaches have been proposed to compare the ability of different models to describe real-world networks and to optimize model parameters to best fit experimental networks. These methods are designed to assign the same probability to networks satisfying the same set of constraint but hardly can take into account the whole network structure [6].

Recently, an information-theoretic framework inspired by quantum statistical mechanics principles has been proposed as a tool to assess and optimize network models [7]. This approach relies on the minimization of the relative entropy based on the network spectral properties. Importantly, this relative entropy does not depend on a distribution of specific descriptors but considers the network as a whole. However, this representation introduces an external, tunable hyperparameter β ; the optimal estimate from the relative entropy minimization procedure critically depends on the choice of β , a major limitation to the practical use of this framework.

Relative entropy is a central concept in thermodynamics of information (see review [8]) and is defined on the basis of the density matrix. In light of this thermodynamic analogy, here we build a physical interpretation of this approach to

network optimization and fitting, where β plays the role of an inverse temperature. This provides criteria for a rigorous selection of the optimal β and enables the practical application of relative entropy minimization in the optimal reconstruction of parameters for different network models.

The paper is structured as follows. First, we present the theoretical framework of classical maximum entropy null models as a way to generate maximally random ensemble of networks with given constraints. We move on to briefly discussing how the Erdős-Rényi random graph and the configuration model emerge naturally from these ideas.

Within the settings of spectral entropies, we propose a practical optimization method based on an approximation of the Laplacian spectrum and give a concise closed-form expression for the gradients of relative entropy with respect to the model parameters. We continue discussing a thermodynamic interpretation of the meaning of relative entropy optimization.

We calculate analytically the optimal temperature parameter of the Erdős-Rényi and planted partition models. Furthermore, we generalize this result to more complex models with the help of numerical simulations showing the advantages of the spectral entropy approach with respect to other maximum likelihood methods.

Finally, we demonstrate the use of spectral entropies for the optimization of a generative model of neural connectivity in a real-world data set.

II. MODELS OF COMPLEX NETWORKS

We summarize here a few definitions which are necessary to make this paper self-contained. Let us consider here simple binary undirected graphs $G = (V, E)$ with $|V| = n$ number of nodes and $|E| = m$ number of links.

The adjacency matrix is denoted as $\mathbf{A} = \{a_{ij}\}$ and the (combinatorial) graph Laplacian as $\mathbf{L} = \mathbf{D} - \mathbf{A}$, where \mathbf{D} is the diagonal matrix of the node degrees. Notably, the (combinatorial) Laplacian matrix associated with an undirected graph is a semi-positive-definite matrix, meaning that all its eigenvalues $\lambda_1 \geq \dots \geq \lambda_n = 0$ are positive and real. A random graph model is an ensemble of networks randomly defined in

^{*}Corresponding author: carlo.nicolini@iit.it

the probability space Ω and distributed around some specific network property. For example, in the Gilbert random graph model the probability distribution $\mathcal{P}(G)$ is sharply peaked in $\mathcal{P}(G) = 1/\Omega$ for graphs with n nodes and exactly m edges and is zero otherwise [1]. This distribution is well described in statistical mechanics as the microcanonical ensemble, as it enforces the constraints strictly.

However, given the combinatorial complexity of dealing with the microcanonical description, it is easier to fix the average value of observables of interest rather than working with exact constraints. This approach gives rise to the canonical ensemble of random graphs [1,9]. This type of models has the same role in the study of complex networks as the Boltzmann distribution in classical statistical mechanics; it gives the maximally uninformed prediction of some network properties subject to the imposed constraints.

These ideas can be dated back to Jaynes’s maximum entropy principle [10]. In this sense, the maximally random ensemble of graphs satisfying the imposed topological constraints on average also takes the name of the exponential random graph model (ERGM) [1] or p^* -model in the social sciences [5].

In its simplest implementation, the ERGM results in the Erdős-Rényi [11] model, where the link probability is constant.

On the other hand, if one wants to generate the maximally random network that maintains the desired degree sequence $\{k_i\}$, then the resulting ensemble is called the undirected binary configuration model (UBCM) [9,12,13]. As the degree is an entirely local topological property, it is affected by the intrinsic properties of vertices. For this reason, one can assign a hidden variable $x_i \geq 0$ to each node. Its value acts as a fitness score, which is hypothesized to be proportional to the expected node degree [12]. If two nodes have a high fitness score, then they are more likely to be connected by a link. In this model one can describe the link probability as the normalized product of their scores [9,13,14], resulting in the following expected link probability:

$$p_{ij} = \mathbb{E}[a_{ij}] = \frac{x_i x_j}{1 + x_i x_j}. \tag{1}$$

The values of x_i are obtained by numerical optimization of a specifically designed likelihood function [13,15]. In this framework, the hidden variables x_i are the Lagrange multipliers of the constrained problem that ensures the expected degree $\langle k_i \rangle = \sum_{i \neq j} \langle a_{ij} \rangle$ of the vertex i equals on average its empirical value k_i . Interestingly, this model highlights the fermionic properties of the links, as they are modeled like particles with only two states, namely the link being present or not.

If the network is sufficiently random, then the degree sequence alone can model the higher-order patterns like the clustering coefficient or the average nearest-neighbor degree. However, deviations of other graph-theoretical measures between model and empirical network are indicative of genuine higher-order patterns, like clustering or rich clubs, not simply accountable by the degree sequence alone [14,15].

III. SPECTRAL ENTROPIES FRAMEWORK

A measure of complexity is central to the understanding of differences and similarities between networks and to decode the information that they represent. Supported by the seminal

demonstration that the von Neumann entropy of a properly defined density matrix may be used for network comparison [7], in this paper we address the unsolved problem of inverse temperature selection and show that the result of model fitting strongly depend on it.

The first observation that an appropriately normalized graph Laplacian can be treated as a density matrix of a quantum system is credited to the authors of Ref. [16]. Indeed, the Laplacian spectrum encloses a number of important topological properties of the graph [17–20]. For instance, the multiplicity of the zero eigenvalue corresponds to the number of connected components, the multiplicity of each eigenvalue is related to graph symmetries [19–21], and the concept of expanders and isoperimetric number are connected to the first- and second-largest eigenvalues [22,23]. Moreover, the graph Laplacian appears often in the study of random walkers [24,25], diffusion [26], combinatorics [27], and a large number of other applications [19,27].

After the first demonstration that a graph can be always represented as a uniform mixture of pure density matrices [16], at least two different definitions of quantum density for complex networks have been used [7,28]. Adopting the notation of quantum physics, the von Neumann density matrix ρ is a Hermitian and positive-definite matrix with unitary trace that admits a spectral decomposition as

$$\rho = \sum_{i=1}^n \lambda_i(\rho) |\phi_i\rangle \langle \phi_i| \tag{2}$$

for an orthonormal basis $\{|\phi_i\rangle\}$ and eigenvalues $\lambda_i(\rho)$. Thus, a density matrix can be represented as a convex combination of pure states [28].

The von Neumann entropy of the density operator ρ can be expressed as the Shannon entropy of its eigenvalues [29]:

$$S(\rho) = -\text{Tr}[\rho \log \rho] = -\sum_{i=1}^n \lambda_i(\rho) \log \lambda_i(\rho), \tag{3}$$

where $\log(\cdot)$ is the principal matrix logarithm [30] when the argument is a matrix. The von Neumann entropy of the density matrix is bounded between 0 and $\log n$ [29]. All logarithms are in natural base.

We adopt the quantum statistical mechanics perspective [7], where the von Neumann density matrix ρ of a complex network is built considering a quantum system with Hamiltonian \mathbf{L} in thermal contact with a heat bath at constant temperature $k_B T = 1/\beta$, where k_B is the Boltzmann constant.

In the perspective of Jaynes’s maximum entropy framework [31], the state of maximum uncertainty about the system, constrained by the conditions $\text{Tr}[\rho] = 1$ and $\langle \mathbf{L} \rangle = \text{Tr}[\rho \mathbf{L}]$, is described by the quantum Gibbs-Boltzmann distribution:

$$\rho = \frac{e^{-\beta \mathbf{L}}}{\text{Tr}[e^{-\beta \mathbf{L}}]}, \tag{4}$$

where here $e^{(\cdot)}$ is the matrix exponential when the argument is a matrix and the denominator is the so-called partition function of the system, i.e., the sum over all possible configurations of the ensemble, and is denoted by $Z = \text{Tr}[e^{-\beta \mathbf{L}}]$. Borrowing the terminology of statistical physics, thermal averages of any graph-theoretical measure \mathbf{O} over the ensemble defined by ρ

are obtained as

$$\langle \mathbf{O} \rangle_\rho = \text{Tr}[\mathbf{O}\rho] = \frac{1}{Z} \text{Tr}[\mathbf{O}e^{-\beta\mathbf{L}}]. \quad (5)$$

The choice of this density matrix for complex networks is supported by the observation that previous definitions of entropy [16,28] in graph theory resulted in violation of sub-additivity [7,32], a central property of entropy [29]. The strength of this definition of von Neumann entropy for graphs lies in the possibility to establish a connection between quantum statistical mechanics and the realm of networks. Moreover, this approach closely resembles the one taken in the study of diffusion on networks [25], where β is no longer interpreted as an inverse temperature of the external heat bath but rather as the diffusion time of a random walker [33,34]. This renders the idea that the network properties can be explored at different scales by varying β [25,32,35]. Indeed, other physical properties of physical systems can be determined by the spectral properties of the graph defining the interaction [36–38].

IV. MODEL OPTIMIZATION

The application of the previously introduced concepts from information theory and statistical mechanics to complex networks offers many intriguing possibilities, the most important one being the quantification of the amount of shared information between graphs and model fitting. The relative entropy $S(\rho\|\sigma)$ between two density matrices ρ and σ is a non-negative quantity that measures the expected amount of information lost when σ is used instead of ρ [7,29] and it is defined as

$$S(\rho\|\sigma) = \text{Tr}[\rho(\log \rho - \log \sigma)] \geq 0. \quad (6)$$

From the linearity of the trace operator, it is apparent that Eq. (6) consists of two terms. The first term is the negative value of entropy of the empirical density ρ . The second term $\text{Tr}[\rho \log \sigma]$ can be seen as the log-likelihood ratio between densities ρ and σ [7,29,39]. For this reason we can also express Eq. (6) as

$$S(\rho\|\sigma) = -S(\rho) - \log \mathcal{L}(\rho, \sigma), \quad (7)$$

where $\log \mathcal{L} = \text{Tr}[\rho \log \sigma]$ is the network log-likelihood. Indeed, Eq. (7) can be thought as a measure to quantify the discrepancy between the density matrix ρ of an observed network and the model density matrix.

In these settings, model optimization corresponds to estimate the parameters $\hat{\theta}$ via minimization of the expectation of the relative entropy over all graphs with parameters θ :

$$\hat{\theta} = \underset{\theta}{\text{argmin}} \mathbb{E}[S(\rho\|\sigma(\theta))]. \quad (8)$$

Rigorous calculation of the expected relative entropy requires the knowledge of the spectral properties of the model Laplacian. These can be obtained via application of random matrix theory or by Monte Carlo sampling [40–42]. However, in this case a continuous gradient-based optimization cannot be applied as the relative entropy is no longer differentiable. Other methods, like simulated annealing [43] or stochastic optimization [44,45] can be applied in this case, with substantial computational burden. Here, for simplicity, we use a different approach. As any random graph model depends

on some parameters $\theta = \{\theta_1, \dots, \theta_k\}$, the model Laplacian is a matrix of random variables, where its elements are drawn from some distribution with parameters θ . For example, in the Erdős-Rényi random graph, the Laplacian off-diagonal elements are Bernoulli random variables with expectation $-p$, while the diagonal elements are binomial random variables with expectation $(n-1)p$. Hence, we denote the expectation operator of the model Laplacian $\mathbf{L}(\theta)$ at fixed parameters θ as $\mathbb{E}[\mathbf{L}(\theta)]$.

As discussed in Appendix A, we approximate the expected relative entropy of Eq. (8) with the relative entropy between the observed density and the density of the expected model Laplacian:

$$\mathbb{E}[S(\rho\|\sigma(\theta))] \approx S(\rho\|\sigma(\mathbb{E}[\mathbf{L}(\theta)])). \quad (9)$$

The accuracy of the approximation in Eq. (9) is higher for large dense networks. Finally, as the expected Laplacian $\mathbb{E}[\mathbf{L}]$ depends continuously on the parameters θ , we can use continuous optimization methods based on the analytically computed gradients with components (see Appendix B for detailed calculation):

$$\frac{\partial S(\rho\|\sigma(\mathbb{E}[\mathbf{L}]))}{\partial \theta_k} = \beta \text{Tr} \left[(\rho - \sigma(\mathbb{E}[\mathbf{L}(\theta)])) \frac{\partial \mathbb{E}[\mathbf{L}(\theta)]}{\partial \theta_k} \right]. \quad (10)$$

This last equation is the basis of the following sections and enables application of gradient-based optimization methods.

A. Thermodynamic interpretation

The Klein inequality states that the quantum relative entropy of two density matrices is always non-negative and zero only in the case $\rho = \sigma$ [29]. It is interesting to rework the expression for the relative entropy by making use of thermodynamic quantities [8,46]. For notational clarity here we set $\mathbf{L}_\sigma := \mathbb{E}[\mathbf{L}(\theta)]$, but the following reasoning works for any two Laplacians \mathbf{L}_ρ and \mathbf{L}_σ . We denote the Helmholtz free energies F_ρ, F_σ of the two systems described by densities ρ and σ as

$$F_\rho = -\beta^{-1} \log Z_\rho,$$

$$F_\sigma = -\beta^{-1} \log Z_\sigma,$$

where the partition functions are computed as $Z_\rho = \text{Tr}[e^{-\beta\mathbf{L}_\rho}]$, $Z_\sigma = \text{Tr}[e^{-\beta\mathbf{L}_\sigma}]$. The ensemble averages of the empirical and model Laplacians are $\langle \mathbf{L}_\rho \rangle_\rho = \text{Tr}[\rho \mathbf{L}_\rho]$ and $\langle \mathbf{L}_\sigma \rangle_\rho = \text{Tr}[\rho \mathbf{L}_\sigma]$, where $\langle \cdot \rangle_\rho$ indicates thermal averaging with respect to the canonical distribution pertaining to the Laplacian of the observed network \mathbf{L}_ρ . After rearrangement of the terms, the expression for the relative entropy described in Eq. (6) becomes

$$S(\rho\|\sigma) = \beta[(F_\rho - F_\sigma) - (\langle \mathbf{L}_\rho \rangle_\rho - \langle \mathbf{L}_\sigma \rangle_\rho)] \geq 0. \quad (11)$$

This expression is indeed general for any two density matrices and not only in the settings discussed in Eq. (9).

Clearly, the Klein inequality implies the following condition, also known as Gibbs' inequality in statistical physics [46]:

$$\langle \mathbf{L}_\sigma \rangle_\rho - \langle \mathbf{L}_\rho \rangle_\rho \geq F_\sigma - F_\rho \quad (12)$$

or, equivalently, rearranging the terms:

$$\langle \mathbf{L}_\sigma \rangle_\rho - F_\sigma \geq \langle \mathbf{L}_\rho \rangle_\rho - F_\rho. \quad (13)$$

As the right-hand side of Eq. (13) is independent on the parameters θ , the minimum relative entropy is obtained by direct minimization of the left-hand side of Eq. (13).

This expression has a profound physical interpretation [8,46–48]. Let us consider a system with Hamiltonian $\mathbf{L}_\rho(\theta^*)$ in equilibrium with a heat bath at inverse temperature β . The system is driven abruptly by an external force from a state described by control parameters θ^* to a state described by parameters θ that we describe by the Hamiltonian $\mathbf{L}_\sigma(\theta)$. The work performed by this force is identified with W : Because of the sudden change in internal parameters, the expectation of W has to be taken with respect to the initial state described by ρ . Hence, this transformation is done by injecting an amount of energy $\langle W \rangle_\rho = \text{Tr}[\rho \mathbf{L}_\sigma] - \text{Tr}[\rho \mathbf{L}_\rho]$ into the system. By looking at the left-hand side of Eq. (12), this is exactly the average change in internal energy. By the Gibbs inequality this is at least as large as the increase in free energy $\Delta F = F_\sigma - F_\rho$. The difference $\langle W \rangle_\rho - \Delta F = \beta^{-1} S(\rho \| \sigma) \geq 0$ is due to the irreversible nature of the process, a result of the abrupt energy injection. In this sense, minimization of the relative entropy is equivalent to finding a set of parameters $\hat{\theta}$ such that the work required to bring the system described by ρ to a state σ is as close as possible to the change of free energy.

This interpretation yields another view on the relative entropy as the amount of work that is needed to counterbalance the transformation from ρ to σ . For instance, in modeling highly modular graphs with a random graph with fixed degree sequence, one loses the information encoded in the neighboring relation of nodes within the same module. This information lost is encoded in the relative entropy. A large amount of work is needed to insert or remove links where they are less likely to be according to the model. At least for the case of modular networks, this work can be thought as the amount of energy needed to “separate” the Laplacian eigenvalues related to the modular structure from the bulk, for example.

In a thermodynamical perspective, the role of β in the spectral entropy framework is therefore clearer: When fitting any model derived from the MaxEnt principle such as the Erdős-Rényi or the configuration model to an observed network, the result is unbiased only in the $\beta \rightarrow 0$ limit. It is in the high-temperature limit that the influence of any specific generative rule giving rise to the observed network becomes independent of node pairs and the tendency to distribute links becomes uniform. The resulting network then will result closer in probability to the average realization described by a maximum entropy model, with lower relative entropy between them.

On the other hand, for β tending to zero, the two density matrices ρ and σ tend to \mathbf{I}/n , so any choice of the parameters trivially yields zero relative entropy.

Hence, to guide the optimization towards a minimum relative entropy configuration in the $\beta \rightarrow 0$ limit, one must start with some initial guess of β_0 , isothermally find a local minimum and afterwards decrease β . Eventually, as β tends to zero, the optimal solution will change slowly while making the relative entropy as small as possible.

Differently from previous network models, the power of the spectral entropy approach is in the ability to assign different probabilities to networks with the same generative parameters, as the whole spectrum is compared, beyond the value of sufficient statistics. It is for this reason that in order to precisely reconstruct the parameters $\tilde{\theta} = \theta^*$, minimization of relative entropy averaged over the whole set of samples is needed.

Finally, we note that the inverse temperature β can also be interpreted as a diffusion time, enabling one to compare two networks at different scales [7,32]. With this analogy in mind, it is clear that diffusion is faster over densely connected sub-graphs and slower over larger and sparser structures. Networks with very strong community structure display a fast diffusion over the modules, resulting in a large change of $S(\rho)$ over a small range of β . Intuitively, in the limit $\beta \rightarrow 0$, diffusion is limited and only conveys information about degrees, a first-order property that is linearly dependent on the adjacency matrix. Hence, we expect that exponential random graphs modeling properties like the total number of links (ER) or the degrees (configuration model), are least biased in the $\beta \rightarrow 0$ limit.

In the following, we show analytically and numerically that correct reconstruction of the empirical edge density parameters is possible only in the limit $\beta \rightarrow 0$ for the Erdős-Rényi and the planted partition model.

B. Erdős-Rényi random graph

The Erdős-Rényi random graph is the simplest example of random graph model [11]. Each pair of nodes is connected by a link with constant probability p . Hence, the expected adjacency matrix and the Laplacian can be written as $\mathbb{E}[\mathbf{A}_\sigma] = p(\mathbf{1} - \mathbf{I})$ and $\mathbb{E}[\mathbf{L}_\sigma] = (n - 1)p\mathbf{I} - p(\mathbf{1} - \mathbf{I})$, where $\mathbf{1}$ is the $n \times n$ matrix of ones and \mathbf{I} is the identity matrix. In this case it is possible to analytically find the optimum solution for the problem of relative entropy minimization. The partition function $Z_\sigma(n, p)$ and the ensemble average of the expected Laplacian $\text{Tr}[\rho \mathbb{E}[\mathbf{L}]]$ of the model are

$$Z_\sigma(p) = (n - 1)e^{-n\beta p} + 1, \quad (14)$$

$$\text{Tr}[\mathbb{E}[\mathbf{L}(p)]\rho] = p[n - R(n, \beta)],$$

where $R(n, \beta) = \text{Tr}[\mathbf{1}\rho] = \sum_{i,j} \rho_{ij}$ is the grand sum of density matrix. Both Z_ρ and F_ρ are observation dependent quantities and must be evaluated numerically from the observed network. Finding the minimum of the left-hand side of Eq. (13) corresponds to setting to zero its derivative with respect to p :

$$\begin{aligned} \frac{\partial}{\partial p} \left[\text{Tr}[\rho \mathbb{E}[\mathbf{L}(p)]] + \frac{\log Z_\sigma(p)}{\beta} \right] &= 0 \\ &= \text{Tr} \left[\frac{\partial \mathbb{E}[\mathbf{L}(p)]}{\partial p} \rho \right] - \frac{n(n - 1)}{e^{\beta n p} + (n - 1)} = 0. \end{aligned} \quad (15)$$

Solving for p , we can find an analytical expression for the optimal density \tilde{p} that can be reconstructed by the model:

$$\tilde{p} = \frac{1}{n\beta} \log \left\{ \frac{R(n, \beta)(n - 1)}{[n - R(n, \beta)]} \right\}. \quad (16)$$

The reconstruction of the observed empirical density $p^* = 2m^*/[n(n - 1)]$ is only possible in the limit $\beta \rightarrow 0$. It can be

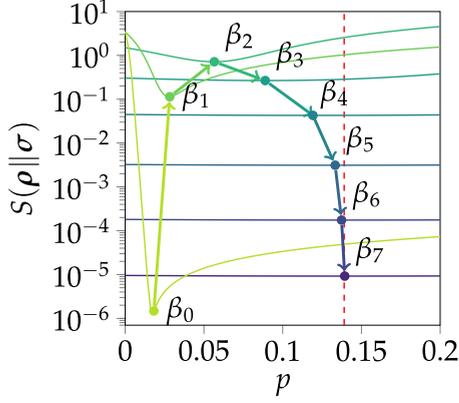


FIG. 1. Relative entropy as function of β and p for the Erdős-Rényi model. The empirical density ρ is obtained from the Laplacian of the karate club graph. The filled dots are the solutions \tilde{p} from Eq. (16) indicating the value of p where minimum relative entropy is attained, in perfect agreement with numerical minima (solid lines). Color encodes decreasing β values, from large (yellow, light gray) β to small β (blue, dark gray). The red vertical dashed line corresponds to the empirical edge density of the karate club graph.

shown, with the help of computer algebra system, that

$$\lim_{\beta \rightarrow 0} \tilde{p} = p^* = \frac{2m^*}{n(n-1)}. \quad (17)$$

A numerical demonstration of this finding is shown in Fig. 1, where the location of the minimum relative entropy as a function of p is shown at different values of β for the density matrix pertaining to the Laplacian of the karate club graph [49]. We also extended the same calculations to the planted-partition model, the simplest extension of the Erdős-Rényi model to networks presenting a community structure [50].

C. Planted partition model

In the planted partition model, the nodal block membership vector $c_i \in \{0, \dots, b\}$ specifies to which one of the b blocks the node i belongs and it has the role of a hyperparameter. The model parameters are the intrablock and interblock link densities p_{in} and p_{out} . The expected adjacency matrix $\mathbb{E}[\mathbf{A}]$ and Laplacian $\mathbb{E}[\mathbf{L}]$ are

$$\mathbb{E}[\mathbf{A}] = \delta p_{\text{in}} + (\mathbf{1} - \delta) p_{\text{out}}, \quad (18)$$

$$\begin{aligned} \mathbb{E}[\mathbf{L}] = & \mathbf{I}[p_{\text{in}}(n/b - 1) + p_{\text{out}}(n/b)(b - 1)] \\ & - \delta p_{\text{in}} - (\mathbf{1} - \delta) p_{\text{out}}, \end{aligned} \quad (19)$$

where $\delta = \{\delta_{c_i, c_j}\}$ is the block assignment matrix. Analogous calculations as in the Erdős-Rényi case can be analytically performed in the planted partition model with exactly b blocks of the same size. In this case, the partition function of the model $Z_\sigma(p_{\text{in}}, p_{\text{out}})$ becomes

$$\begin{aligned} Z_\sigma(p_{\text{in}}, p_{\text{out}}) &= \text{Tr}[e^{-\beta \mathbb{E}[\mathbf{L}_\sigma]}] \\ &= (n - b) \exp[-\beta n(p_{\text{in}} + (b - 1)p_{\text{out}})/b] \\ &\quad + (b - 1) \exp[-\beta n p_{\text{out}}] + 1. \end{aligned} \quad (20)$$

Setting the gradients of relative entropy (10) to zero results in a system of two equations:

$$\begin{aligned} \text{Tr} \left[\frac{\partial \mathbb{E}[\mathbf{L}_\sigma(p_{\text{in}}, p_{\text{out}})]}{\partial p_{\text{in}}} \rho \right] + \frac{1}{\beta} \frac{\partial \log Z_\sigma(p_{\text{in}}, p_{\text{out}})}{\partial p_{\text{in}}} &= 0, \\ \text{Tr} \left[\frac{\partial \mathbb{E}[\mathbf{L}_\sigma(p_{\text{in}}, p_{\text{out}})]}{\partial p_{\text{out}}} \rho \right] + \frac{1}{\beta} \frac{\partial \log Z_\sigma(p_{\text{in}}, p_{\text{out}})}{\partial p_{\text{out}}} &= 0. \end{aligned} \quad (21)$$

An analytical solution is possible for $b = 2$ blocks:

$$\tilde{p}_{\text{in}} = \frac{1}{\beta n} \log \left[\frac{(n - 2)^2 R (2Q - R)}{(n - 2Q)^2} \right], \quad (22)$$

$$\tilde{p}_{\text{out}} = \frac{1}{\beta n} \log \left[\frac{R}{2Q - R} \right], \quad (23)$$

where R is the grand-sum of ρ and $Q = \text{Tr}[\delta \rho]$. As in the previous case, the empirical intrablock and interblock densities p_{in}^* and p_{out}^* can be reconstructed only in the limit of infinite temperature:

$$\begin{aligned} \lim_{\beta \rightarrow 0} \tilde{p}_{\text{in}} &= p_{\text{in}}^*, \\ \lim_{\beta \rightarrow 0} \tilde{p}_{\text{out}} &= p_{\text{out}}^*. \end{aligned} \quad (24)$$

Unfortunately, though, the calculations performed for these two last examples cannot be straightforwardly extended to the configuration model (UBCM) and other more complex variants of the exponential random graph model, as the expression of the partition function Z_σ for general models is too complex for a fully analytical treatment. Therefore, we rely on numerical simulations to show that the limit $\beta \rightarrow 0$ yields a correct reconstruction of model parameters also for the UBCM.

D. Configuration model

In the UBCM the model parameters θ are the n hidden variables $\mathbf{x} = \{x_i\}$. Given some network with fixed degree sequence, we can consider its Laplacian \mathbf{L}_ρ as the Laplacian of a graph sampled from the UBCM ensemble.

Importantly, as \mathbf{L}_ρ is a realization of a random graph ensemble, we cannot expect to be able to reconstruct exactly the parameters \mathbf{x}^* which generated that network. As a demonstration of this concept we generated a random network with $n = 40$ nodes and a degree sequence sampled from the uniform distribution.

In order to estimate the variables \mathbf{x} in the limit $\beta \rightarrow 0$ we adopted an iterative scheme of temperature increase and equilibration, detailed in Fig. 1. Starting with the system in cold state (high value of β) and a random configuration of the variables \mathbf{x} with $x_i \geq 0$, we let the two systems equilibrate by looking for the configuration \mathbf{x} that set the gradients in Eq. (10) to zero (minimum relative entropy) to a numeric error threshold ϵ . We then decrease β by a constant factor γ and repeat the above procedure starting from the previously computed optimal configuration until convergence.

As a measure of convergence, we chose the difference of the total number of links, which in a thermodynamic interpretation equals to half the difference of the total energies: $\Delta m = (\text{Tr}[\mathbb{E}[\mathbf{L}_\sigma]] - \text{Tr}[\mathbf{L}_\rho])/2$. The bigger the absolute value of

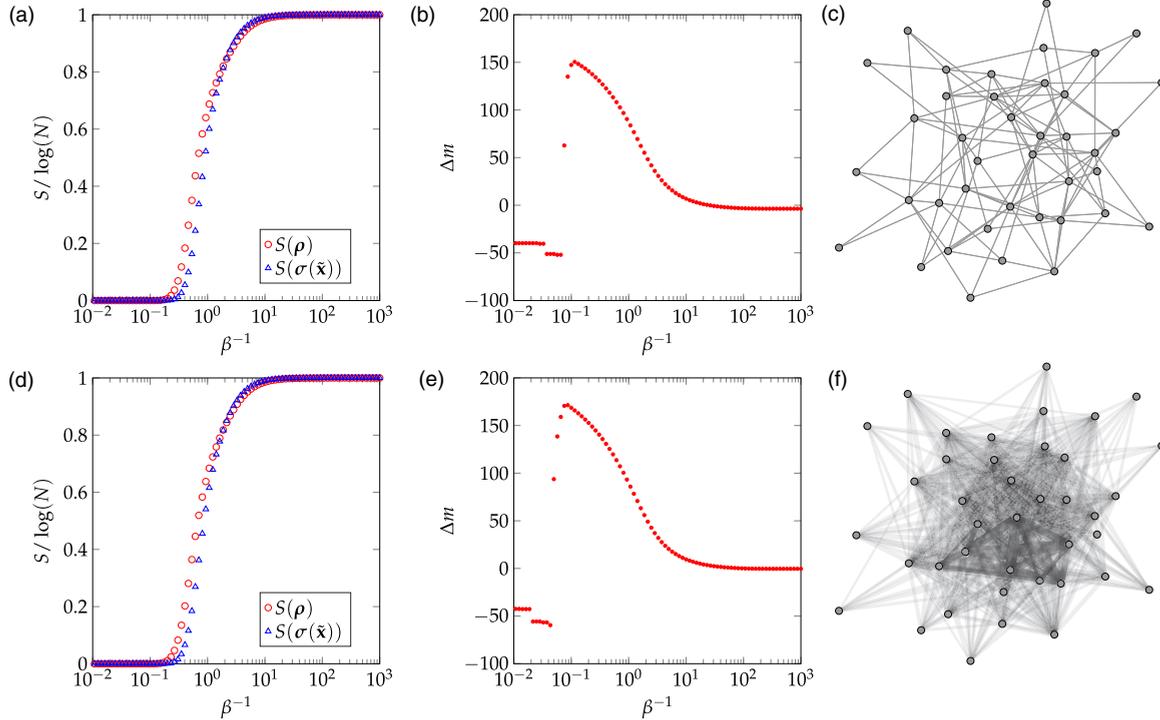


FIG. 2. Normalized von Neumann entropy and reconstruction error of a random network (top panels) and its ensemble average with the same degree sequence (bottom panels). In panels (a) and (d) the red circles are the spectral entropies of the observed network described by density ρ , and blue triangles are the spectral entropies of the model network at optimum parameters $\sigma(\bar{\theta})$. In panel (b) a systematic error exists at every β (even in the limit $\beta \rightarrow 0$): in the spectral framework the UBCM cannot fit all the specific properties of the highly ordered network. On the other hand, an ensemble average network has a structure that can be better reconstructed by the UBCM, as shown by the error Δm going to zero in panel (e).

Δm , the less similar the empirical network is from the average realization of the ensemble.

In Fig. 2(a) we plotted the spectral entropies of the empirical network (red line) and the fitted model at optimal solution $\bar{\mathbf{x}}$ (blue line) as a function of β . Fig. 2(b) shows the difference in the number of links Δm as a function of β . At the optimal solution for $\beta \rightarrow 0$, there is a small deviation in total number of links ($\Delta m \approx -3.64$). However, the deviation of the reconstructed parameters $\bar{\mathbf{x}}$ from \mathbf{x}^* can be reduced with enough samples of the random graph ensemble, as shown in Figs. 2(d)–2(f).

As a second example we chose a toy network consisting of a number of cliques of increasing size connected in a ring, and one of its degree-preserving random rewirings [Figs. 3(c) and 3(f)]. In the ordered case, the clique structure cannot be accounted by a first-order average model alone, making that specific instance highly unlikely in the spectral entropies framework when sampling from the configuration model. Therefore, following the fitting procedure described above, one can see a significant difference in the number of links between model and data Δm even at a very small beta [Fig. 3(b)]. In other words, in the ordered case, the degree sequence alone cannot explain the differences in the spectral entropies, thus indicating the presence of genuine regular patterns that substantially alter the properties of diffusion of the random walker defined by the density matrix. Indeed, this difference reflects the intrinsic inability of the model to account for the characteristic structure of the underlying network.

After fitting, nonvanishing Δm serves as an indicator of the presence of ordered patterns in the given network that are not explained by this model alone. To test this idea we applied the same optimization technique to the degree-preserving randomly rewired network in Fig. 3(f) and plotted the results in Figs. 3(d) and 3(e). In this case the random rewiring made the empirical network more adherent to the optimal reconstruction by the model and the difference in the total number of links at β close to zero is much smaller than in the ordered case. This is also evident by the higher similarity of the spectral entropies, as shown in Fig. 3(d).

Importantly, the spectral entropy optimization framework described above can be applied to descriptive network models other than those described by the exponential random graph model.

V. SPATIAL MODELS

The embedding of a network in a two- or three-dimensional space has bearings on its topological properties. When the formation of links has a cost associated with distance, the model must accommodate additional spatial constraints, which introduce correlation between topological and geometrical organization [51]. An example is represented by neural networks, in which communication between neurons implies a metabolic cost that depends on their distance [52,53]. The material and metabolic constraints of neuronal wiring are factors that contributed to shaping brain architecture [52,54,55].

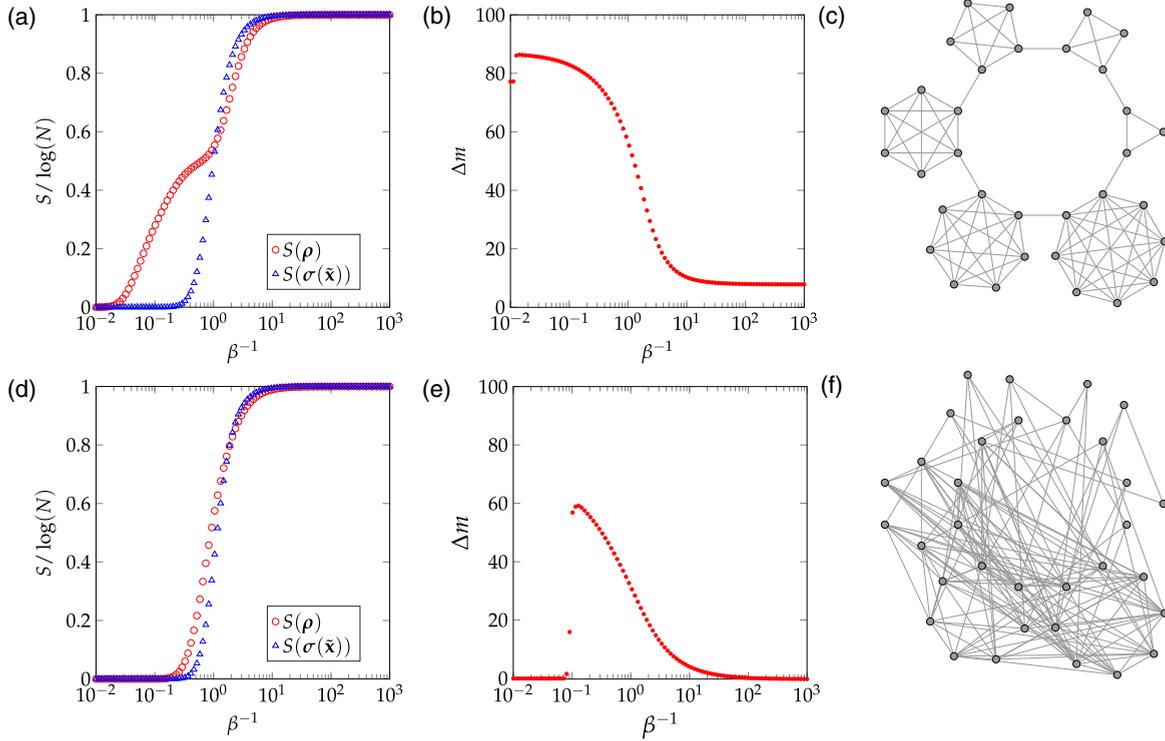


FIG. 3. Normalized von Neumann entropy and reconstruction error of two networks: a ring of cliques (top panels) and its degree-preserving random rewiring (bottom panels). In panels (a) and (d) the red circles are the spectral entropies of the observed network described by density ρ , blue triangles are the spectral entropies of the model network at optimum parameters $\sigma(\hat{\theta})$. In panel (b) a systematic error exists at every β (even in the limit $\beta \rightarrow 0$): The UBCM cannot fit all the specific properties of the highly ordered network. On the other hand, a degree preserving rewired network has a structure that can be better reconstructed by the UBCM, as shown by the error Δm going to zero in panel (e).

Computational and empirical studies converged on the result that a multiscale organization of modules inside modules is the one that satisfies the constraints imposed by minimization of energetic cost and spatial embedding [52,53,56,57]. Here wiring cost includes the physical volume of axons and synapses, the energetic demand for signal transmission, additional processing cost for noise correction over long-distance signaling, and sustenance of the necessary neuroglia that support neuronal activities [52].

Therefore, it is tempting to assume that the expected number of neural fibers between two areas could be expressed as a

decreasing function of their length. With this hypothesis in mind, we verified the ability of our optimization approach to work with a simple descriptive model of the observed neural connectivity in the macaque cortex [58,59]. The model, called exponential distance rule (EDR), is a dense weighted network model describing the decline in the expected number of axonal projections w_{ij} as a function of the interareal spatial distances d_{ij} and a tunable decay parameter $\ell \in \mathbb{R}$:

$$\mathbb{E}[w_{ij}] = C e^{-\ell d_{ij}}, \tag{25}$$

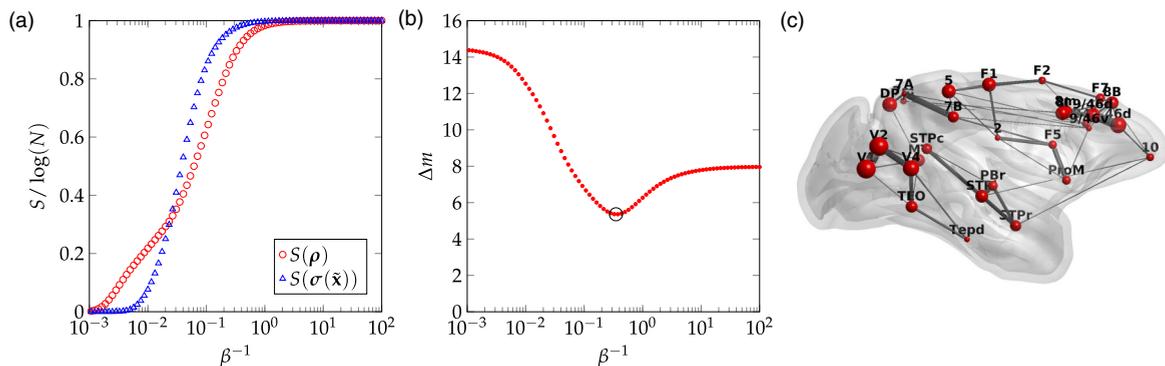


FIG. 4. Normalized von Neumann entropy and reconstruction error of the macaque brain connectivity network in panel (c). In panel (a) the red circles are the spectral entropies of the observed network described by density ρ , and blue triangles are the spectral entropies of the model network at optimum parameter $\sigma(\hat{\theta})$. The optimal hyperparameter β is found in panel (b) where the reconstruction error Δm achieve its minimum denoted by a black circle.

where C is a normalization constant. Here the distances d_{ij} are measured along the shortest path connecting areas via white matter, approximating the axonal distance [58].

We used a data set of corticocortical connectivity generated from retrograde tracing experiments in the macaque brain [58,59]. Following the procedure introduced in the previous sections, we fitted the macaque connectome network with the EDR model. Differently from the random graph models described in previous sections, we found an optimal inverse-temperature parameter β that minimizes the reconstruction error, as shown in Fig. 4(b). At this optimal $\beta^* \approx 3.05$ the reconstructed decay parameter $\tilde{\ell} = 0.1505 \text{ mm}^{-1}$ is comparable to the values obtained by the three methods applied in the original paper from Ref. [59]. The nonvanishing difference in total weights between reconstructed and object networks indicates that the model cannot account completely for the structures such as the high-density core observed in the connectome [59]. Hence, a nonzero optimal value for β suggests the existence of a scale at which the model best describes the topological properties of the network.

VI. CONCLUSION

The spectral entropies framework enables comparing networks taking into account the whole structure at multiple scales. However, this approach introduces a hyperparameter β that plays the role of an inverse temperature and whose tuning is critical for the correct estimate of the model parameters.

Leveraging a thermodynamic analogy, we have shown that the optimal value of the hyperparameter is model dependent and reflects the scales at which the model best describes the empirical network. Moreover, we have described procedures to determine β for the model parameter optimization and for a tractable approximation of the expected relative entropy.

Specifically, we focused on three examples from the exponential random graph model, namely the Erdős-Rényi, a planted partition and undirected binary configuration model. In the Erdős-Rényi model and in the planted partition model, we analytically demonstrated that correct reconstruction is possible only in the infinite temperature limit. In the configuration model this hypothesis was verified numerically. The presence of ordered structures unaccounted for by the model is reflected in a bias of the total energy, corresponding to the total number of links in the reconstructed network.

Motivated by these findings in synthetic networks, we applied the spectral entropy framework to a real-world network of the macaque brain structural connectome. A structural connectome is a spatial network whose development is thought to be constrained by geometrical and wiring cost factors. Hence, we evaluated an exponential distance rule model that assumes that the weight of interareal connections is a decreasing function of distance. We demonstrate the existence of a nonzero optimal value of β for the computation of model parameters. However, the residual energy bias indicates the network structure at certain scales cannot be described by the exponential distance rule model alone.

The procedures demonstrated here make it possible to use relative entropy methods for practical applications to the

study of models of real-world networks, effectively realizing a conceptual step from classical maximum likelihood methods to their density matrix based counterparts.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant No. 668863.

C.N. and V.V. contributed equally to this work.

APPENDIX A: APPROXIMATION OF EXPECTED RELATIVE ENTROPY

We can exploit the commutativity and linearity of the trace and expectation operators to obtain a simpler expression for the expected relative entropy:

$$\mathbb{E}[S(\rho \parallel \sigma(\theta))] = \text{Tr}[\rho \log \rho] - \text{Tr}[\rho \mathbb{E}[\log \sigma(\theta)]]. \quad (\text{A1})$$

By the positive definiteness of the density σ , we have

$$\log(\sigma(\theta)) = -\beta \mathbf{L}(\theta) - \mathbf{I} \log Z(\theta). \quad (\text{A2})$$

Plugging this into the expression of the expected relative entropy we obtain

$$\mathbb{E}[S(\rho \parallel \sigma(\theta))] = \text{Tr}[\rho(\log \rho + \beta \mathbb{E}[\mathbf{L}(\theta)] + \mathbf{I} \mathbb{E}[\log Z(\theta)])]. \quad (\text{A3})$$

This last expression depends on the expected Laplacian of the model $\mathbb{E}[\mathbf{L}(\theta)]$ and on the expected log-partition function $\mathbb{E}[\log Z(\theta)]$. An analytical estimate of the expected Laplacian as function of the parameters θ can readily be obtained, but computation of the expected log-partition function $\mathbb{E}[\log Z(\theta)]$ is more difficult and requires techniques from random matrix theory [40–42]. This is clear as the trace of a matrix is equal to the sum of its eigenvalues, yielding:

$$\mathbb{E}[\log Z(\theta)] = \mathbb{E} \left[\log \left(\sum_{i=1}^n e^{-\beta \lambda_i(\mathbf{L}(\theta))} \right) \right]. \quad (\text{A4})$$

We can estimate the expected log-partition function by means of matrix concentration arguments [60,61]. A random matrix is said to *concentrate* when, given some spectral norm, one can tightly bound the spectral norm of the difference from its expected value [62]. In the case of network-related matrices, the eigenvalues of the Laplacian and those of its expectation are strictly related and can be tightly bounded with high probability [61,63–65]. This approximation becomes more precise, the larger and denser the graphs are [63,66,67], an effect of the concentration of measure phenomenon. Therefore, following the ideas presented in Refs. [61,63,64] that apply in our same settings, we replace $\lambda_i[\mathbf{L}(\theta)]$ with their counterparts from the expected Laplacian $\lambda_i(\mathbb{E}[\mathbf{L}])$. Substituting back, we recover an expression that involves the relative entropy between the observed density and the density of the expected Laplacian:

$$\mathbb{E}[S(\rho \parallel \sigma(\theta))] \approx S(\rho \parallel \sigma(\mathbb{E}[\mathbf{L}])). \quad (\text{A5})$$

APPENDIX B: GRADIENTS OF THE RELATIVE ENTROPY

Here we present the analytical calculation of the gradients of the relative entropy described in Eq. (10). We can decompose the relative entropy using Eq. (A2) as

$$S(\rho \|\sigma(\mathbb{E}[\mathbf{L}(\theta)])) = \beta \text{Tr}[\rho \mathbb{E}[\mathbf{L}(\theta)]] + \log \text{Tr}[e^{-\beta \mathbb{E}[\mathbf{L}(\theta)]}], \quad (\text{B1})$$

where we have used the fact that $\text{Tr}[\rho \mathbf{I}] = \text{Tr}[\rho] = 1$ by definition of density matrix. Taking the derivatives with respect to the parameter θ_k , and by linearity of the trace operator, we get:

$$\frac{\partial S(\rho \|\sigma(\mathbb{E}[\mathbf{L}(\theta)]))}{\partial \theta_k} = \beta \text{Tr} \left[\rho \frac{\partial \mathbb{E}[\mathbf{L}(\theta)]}{\partial \theta_k} \right] + \frac{\partial}{\partial \theta_k} \log \text{Tr}[e^{-\beta \mathbb{E}[\mathbf{L}(\theta)]}]. \quad (\text{B2})$$

The following identity holds for the derivatives of the exponential map:

$$\frac{d}{dt} \text{Tr}[e^{\mathbf{X}(t)}] = \text{Tr} \left[e^{\mathbf{X}(t)} \frac{d\mathbf{X}(t)}{dt} \right], \quad (\text{B3})$$

so we can simply compute the second term involving the log-trace by standard calculus tools. After some algebraic manipulation, we finally arrive to the expression for the

derivative of the relative entropy with respect to the model parameters as described in the main text:

$$\frac{\partial S(\rho \|\sigma(\mathbb{E}[\mathbf{L}(\theta)]))}{\partial \theta_k} = \beta \text{Tr} \left[(\rho - \sigma(\mathbb{E}[\mathbf{L}(\theta)])) \frac{\partial \mathbb{E}[\mathbf{L}(\theta)]}{\partial \theta_k} \right]. \quad (\text{B4})$$

APPENDIX C: MODEL OPTIMIZATION

The fitting problem scales both with the number of nodes and with the number of free parameters of the model. Here we give a short overview of the main computation bottlenecks of the optimization method. For every configuration of the parameters θ , obtaining the relative entropy involves getting the eigenvalues of the model Laplacian and computing traces of the matrix product between ρ and \mathbf{L}_σ . This whole operation scales at worst as $O(N^\gamma)$, with $2 < \gamma < 2.4$. At each point then, computation of the gradients requires another $O(kN^2)$ operations, where k is the number of parameters. When minimization is implemented as a variant of the gradient descent method, the average time complexity to reach convergence is in the order of $O(N^2)$. Hence, the total time of model fitting is $O(kN^{2+\gamma})$ for fixed β . In Algorithm I the detailed pseudocode of the optimization procedure used in this work is given. The lines 5 to 9 are a quasi-Newton optimization method [68], with h being a subroutine of step-size tuning called ‘‘line-search’’ in the optimization literature. Outside this inner loop is the temperature increase scheme, iterated until β is very close to 0.

Algorithm 1 Relative entropy optimization

<pre> 1: $\beta \leftarrow \beta_0$ 2: $\theta_{t,\beta} \leftarrow \theta_{t_0,\beta_0}$ 3: repeat 4: $t \leftarrow 1$ 5: repeat 6: $\eta_t \leftarrow h(S, \eta_t)$ 7: $\theta_{t,\beta} \leftarrow \theta_{t,\beta} - \eta_t \nabla_\theta S_\beta(\rho \ \sigma(\theta))$ 8: $t \leftarrow t + 1$ 9: until $\ \nabla_\theta S_\beta(\rho \ \sigma(\theta))\ \leq \epsilon$ 10: $\hat{\theta}_\beta \leftarrow \theta_{t,\beta}$ 11: $\beta \leftarrow \beta/\gamma$ 12: until β is sufficiently small </pre>	<pre> ▷ Initialize β to large initial value. ▷ Initialize random configuration ▷ Temperature increase schedule ▷ Update counter ▷ Isothermal optimization loop ▷ Line-search routine h ▷ Update ▷ Check convergence ▷ Store optimal configuration ▷ Decrease β (increase temperature) </pre>
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

[1] M. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, 2010).
[2] A.-L. Barabasi and R. Albert, *Science* **286**, 509 (1999).
[3] G. Caldarelli, *Scale-Free Networks: Complex Webs in Nature and Technology* (Oxford University Press, Oxford, 2007).
[4] E. Bullmore and O. Sporns, *Nat. Rev. Neurosci.* **10**, 186 (2009).
[5] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications* (Cambridge University Press, Cambridge, 1994), Vol. 8.

[6] T. Squartini and D. Garlaschelli, *Maximum-Entropy Networks: Pattern Detection, Network Reconstruction and Graph Combinatorics* (Springer International Publishing, 2017).
[7] M. De Domenico and J. Biamonte, *Phys. Rev. X* **6**, 041062 (2016).
[8] J. M. Parrondo, J. M. Horowitz, and T. Sagawa, *Nat. Phys.* **11**, 131 (2015).
[9] J. Park and M. E. J. Newman, *Phys. Rev. E* **70**, 066117 (2004).

- [10] E. T. Jaynes, *Phys. Rev.* **106**, 620 (1957).
- [11] P. Erdős and A. Rényi, *Publ. Math.* **6**, 290 (1959).
- [12] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Muñoz, *Phys. Rev. Lett.* **89**, 258702 (2002).
- [13] T. Squartini, R. Mastrandrea, and D. Garlaschelli, *New J. Phys.* **17**, 023052 (2015).
- [14] T. Squartini and D. Garlaschelli, in *Econophysics of Agent-Based Models*, edited by F. Abergel, H. Aoyama, B. K. Chakrabarti, A. Chakraborti, and A. Ghosh (Springer International Publishing, Cham, 2014) pp. 161–186.
- [15] D. Garlaschelli and M. I. Loffredo, *Phys. Rev. E* **78**, 015101 (2008).
- [16] S. L. Braunstein, S. Ghosh, and S. Severini, *Ann. Comb.* **10**, 291 (2006).
- [17] E. Estrada, *The Structure of Complex Networks: Theory and Applications* (Oxford University Press, Inc., New York, 2011).
- [18] W. N. Anderson, *Linear Multilinear A.* **18**, 141 (1985).
- [19] R. Merris, *Linear Algebra Appl.* **197–198**, 143 (1994).
- [20] S. C. de Lange, M. A. de Reus, and M. P. van den Heuvel, *Front. Comput. Neurosci.* **7**, 1 (2014).
- [21] S. C. de Lange, M. P. van den Heuvel, and M. A. de Reus, *NeuroImage* **141**, 357 (2016).
- [22] J. Cheeger, in *Problems in Analysis (Papers dedicated to Salomon Bochner, 1969)*, edited by G. R. C. (Princeton Univ. Press, Princeton, NJ, 1970), pp. 195–199
- [23] L. Donetti, F. Neri, and M. A. Muñoz, *J. Stat. Mech-Theory E* (2006) P08007.
- [24] L. Lovász, *Bolyai Math. Stud.* **2**, 1 (1993).
- [25] N. Masuda, M. A. Porter, and R. Lambiotte, *Phys. Rep.* **716**, 1 (2017).
- [26] A. J. Bray and G. J. Rodgers, *Phys. Rev. B* **38**, 11461 (1988).
- [27] B. Mohar, in *Graph Theory, Combinatorics, and Applications* (Wiley, 1991), pp. 871–898.
- [28] K. Anand, G. Bianconi, and S. Severini, *Phys. Rev. E* **83**, 036109 (2011).
- [29] M. M. Wilde, *Quantum Information Theory* (Cambridge University Press, Cambridge, 2013).
- [30] N. J. Higham, *Functions of Matrices: Theory and Computation* (SIAM, Philadelphia, PA, 2008).
- [31] E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, Cambridge, 2003).
- [32] J. Biamonte, M. Faccin, and M. De Domenico, *arXiv:1702.08459* (2017).
- [33] E. Estrada and N. Hatano, *Phys. Rev. E* **77**, 036111 (2008).
- [34] M. Faccin, T. Johnson, J. Biamonte, S. Kais, and P. Międał, *Phys. Rev. X* **3**, 041007 (2013).
- [35] E. Estrada, N. Hatano, and M. Benzi, *Phys. Rep.* **514**, 89 (2012).
- [36] R. Burioni, D. Cassi, and A. Vezzani, *Phys. Rev. E* **60**, 1500 (1999).
- [37] R. Burioni, D. Cassi, and C. Destri, *Phys. Rev. Lett.* **85**, 1496 (2000).
- [38] M. I. Berganza and L. Leuzzi, *Phys. Rev. B* **88**, 144104 (2013).
- [39] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley-Interscience, New York, 2006).
- [40] R. R. Nadakuditi and M. E. J. Newman, *Phys. Rev. Lett.* **108**, 188701 (2012).
- [41] T. P. Peixoto, *Phys. Rev. Lett.* **111**, 098701 (2013).
- [42] R. R. Nadakuditi and M. E. J. Newman, *Phys. Rev. E* **87**, 012803 (2013).
- [43] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, *Science* **220**, 671 (1983).
- [44] H. Robbins and S. Monro, *Ann. Math. Statist.* **22**, 400 (1951).
- [45] J. Kiefer and J. Wolfowitz, *Ann. Math. Statist.* **23**, 462 (1952).
- [46] N. Merhav, *Found. Trends Commun. Inf. Theory* **6**, 1 (2010).
- [47] N. Merhav, *IEEE Trans. Inf. Theory* **56**, 4274 (2010).
- [48] S. Deffner and E. Lutz, *Phys. Rev. Lett.* **105**, 170402 (2010).
- [49] W. W. Zachary, *J. Anthropol. Res.* **33**, 452 (1977).
- [50] A. Condon and R. M. Karp, *Randomization, Approximation, and Combinatorial Optimization. Algorithms and Techniques*, edited by D. S. Hochbaum, K. Jansen, J. D. P. Rolim, and A. Sinclair, Lecture Notes in Computer Science, Vol. 1671 (Springer, Berlin, Heidelberg, 1999), pp. 221–232.
- [51] M. Barthélemy, *Phys. Rep.* **499**, 1 (2011).
- [52] E. Bullmore and O. Sporns, *Nat. Rev. Neurosci.* **13**, 336 (2012).
- [53] R. F. Betzel and D. S. Bassett, *J. R. Soc. Interface* **14**, 20170623 (2017).
- [54] C. J. Stam and E. C. W. van Straaten, *Clin. Neurophysiol.* **123**, 1067 (2012).
- [55] C. Ribault, K. Sekimoto, and A. Triller, *Nat. Rev. Neurosci.* **12**, 375 (2011).
- [56] G. Doucet, M. Naveau, L. Petit, N. Delcroix, L. Zago, F. Crivello, G. Jobard, N. Tzourio-Mazoyer, B. Mazoyer, E. Mellet, and M. Joliot, *J. Neurophysiol.* **105**, 2753 (2011).
- [57] M. Kaiser and C. C. Hilgetag, *PLoS Comput. Biol.* **2**, e95 (2006).
- [58] N. T. Markov *et al.*, *Cereb. Cortex* **24**, 17 (2014).
- [59] M. Ercsey-Ravasz, N. T. Markov, C. Lamy, D. C. VanEssen, K. Knoblauch, Z. Toroczka, and H. Kennedy, *Neuron* **80**, 184 (2013).
- [60] J. A. Tropp, *Found. Trends Mach. Learn.* **8**, 1 (2015).
- [61] J. Cape, M. Tang, and C. E. Priebe, *Electron. J. Stat.* **11**, 3954 (2017).
- [62] R. Qiu and M. Wicks, *Cognitive Networked Sensing and Big Data* (Springer, New York, 2014).
- [63] F. Chung and M. Radcliffe, *Electron. J. Combinat.* **18**, 215 (2001).
- [64] R. I. Oliveira, *arXiv:0911.0600*.
- [65] V. M. Preciado and M. A. Rahimian, *IEEE Trans. Netw. Sci. Eng.* **4**, 215 (2017).
- [66] C. M. Le, E. Levina, and R. Vershynin, *arXiv:1502.03049*.
- [67] C. M. Le, E. Levina, and R. Vershynin, *Random Struct. Algor.* **51**, 538 (2017).
- [68] J. Nocedal and S. Wright, in *Numerical Optimization* (Springer, New York, 2006), pp. 101–134.