

# Spectral Entropies as Information-Theoretic Tools for Complex Network Comparison

Manlio De Domenico<sup>1,\*</sup> and Jacob Biamonte<sup>2</sup>

<sup>1</sup>*Departament d'Enginyeria Informàtica i Matemàtiques,  
Universitat Rovira i Virgili, 43007 Tarragona, Spain*

<sup>2</sup>*Quantum Complexity Science Initiative, Department of Physics, University of Malta, Msida MSD 2080,  
Malta and Institute for Quantum Computing, University of Waterloo, Waterloo, N2L 3G1 Ontario, Canada*  
(Received 25 August 2016; revised manuscript received 17 October 2016; published 21 December 2016)

Any physical system can be viewed from the perspective that information is implicitly represented in its state. However, the quantification of this information when it comes to complex networks has remained largely elusive. In this work, we use techniques inspired by quantum statistical mechanics to define an entropy measure for complex networks and to develop a set of information-theoretic tools, based on network spectral properties, such as Rényi  $q$  entropy, generalized Kullback-Leibler and Jensen-Shannon divergences, the latter allowing us to define a natural distance measure between complex networks. First, we show that by minimizing the Kullback-Leibler divergence between an observed network and a parametric network model, inference of model parameter(s) by means of maximum-likelihood estimation can be achieved and model selection can be performed with appropriate information criteria. Second, we show that the information-theoretic metric quantifies the distance between pairs of networks and we can use it, for instance, to cluster the layers of a multilayer system. By applying this framework to networks corresponding to sites of the human microbiome, we perform hierarchical cluster analysis and recover with high accuracy existing community-based associations. Our results imply that spectral-based statistical inference in complex networks results in demonstrably superior performance as well as a conceptual backbone, filling a gap towards a network information theory.

DOI: [10.1103/PhysRevX.6.041062](https://doi.org/10.1103/PhysRevX.6.041062)

Subject Areas: Complex Systems, Quantum Information,  
Statistical Physics

## I. INTRODUCTION

Shannon's entropy [1] and information-theoretic-derived measures have been successfully applied in a range of disciplines, from revealing time scale dependence in neural coding [2,3] to quantifying quantum information [4–6] and the complexity of genetic sequences [7,8] or to unravel the mesoscale organization of interconnected systems [9–12], to cite just a few emblematic achievements. However, when it comes to complex networks, an appropriate definition of entropy has remained elusive—applicability being often limited to the probability distribution of some network descriptor (such as the normalized distribution of node degrees).

Complex network theory dates to the discovery of several fundamental features [13,14], and complex networks have been widely used to model and better understand the organization of complex systems and their dynamics [15–20], the controllability of their constituents [21], and

their resilience to structural and dynamical perturbations [22–27]. A recent and ongoing drive in complex networks is to merge ideas with quantum information science, including entangled networks for communication [28–30], developing a theory of node centrality in quantum walks on graphs [31], and the detection of community structures formed in quantum systems [32].

In the same spirit as the Church-Turing-Deutsch principle [33], like all physical systems it is possible to view complex networks in terms of information processing, in which information changes in time from an input to an output state of a system. This then necessitates a method to quantify this information, its time dependence in terms of storage and transfer of information, ideally between levels of a multilayer system, which can further be imperfect due to randomness or noise. A similar challenge was addressed by quantum theorists beginning decades ago when faced with quantifying informatic properties of quantum states [4–6,34–37]. Indeed, the modern theory of quantum information, as the name suggests, is fundamentally built upon the quantification of physical information, entropic-based quantifications of nonlocality, entanglement [35] and the inherent complexity of the quantum model [37]. These decades of research have placed entropic measures central to the modern theory, leading to quantum generalizations of Shannon's classical information theory [4–6,34–37].

\*Corresponding author.  
manlio.dedomenico@urv.cat

*Published by the American Physical Society under the terms of the Creative Commons Attribution 3.0 License. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.*

However, for a variety of reasons, these results cannot be applied directly to complex networks, although classical information-theoretic tools have been successfully adopted to solve specific problems [11,38–46].

Here, we address a similarly motivated challenge: inspired by how entropy is calculated in quantum systems, we define an interconnectivity-based density matrix to calculate the von Neumann entropy of a network. We analytically prove that our definition satisfies the desired additivity properties similar to quantum thermodynamic entropy—though the proof of this differs tremendously from the quantum case—which opens the door to avoid the shortcomings, recently pointed out in Ref. [47], imposed by subadditivity failing, as found in past approaches.

We exploit this entropy to develop a set of information-theoretic tools that apply to complex networks, such as Rényi  $q$  entropy, generalized Kullback-Leibler and Jensen-Shannon divergences, thus providing a backbone to an information-theoretic approach to network science. Our framework, based on spectral properties of networks, allows one to probe contemporary problems faced in complex network science. For instance, we show that Kullback-Leibler minimization can be used to infer the parameter(s) of a model aimed to fit an observed network. By exploiting results of classical information theory, we are able to introduce the spectral counterpart of likelihood maximization and use it in practical applications to fit network models and perform model selection based on appropriate Akaike and Bayesian information criteria, as well as minimum description length. The strength of our approach relies on the fact that it uses the network as a whole, instead of a subset of the network’s descriptors, to attack parameter inference and model selection problems.

Another important by-product of the proposed information-theoretic framework is the possibility to quantify the distance between complex networks. This problem is of great interest in network science [48] and, more specifically, in recent applications concerning multilayer systems, such as time-varying and multiplex networks [49–51]. This type of system consists of nodes replicated across several networks that exhibit different types of relationships or connectivity [52]. The possibility to compare layers from an information-theoretic point of view has been recently explored to aggregate them in order to reduce the structure and the complexity of biological, transportation, and social multiplex networks [47]. To numerically probe our method, we cluster the layer of an empirical system and compare it against the existing classification.

## II. VON NEUMANN ENTROPY OF A COMPLEX NETWORK

In quantum mechanics, probability distributions are encoded by density matrices. A density matrix  $\rho$  is a Hermitian and positive semidefinite matrix, with trace

equal to unity, which is used to represent both mixed and pure quantum states. A system is in a pure state  $|\psi\rangle$  if and only if the bound  $\text{Tr}\rho^2 \leq 1$  is saturated. The density matrix admits a spectral decomposition as

$$\rho = \sum_{i=1}^N \lambda_i |\phi_i\rangle \langle \phi_i| \quad (1)$$

for an orthonormal basis  $\{|\phi_i\rangle\}$ , where  $\lambda_i$  are non-negative eigenvalues that sum up to 1.

The density matrix allows one to define the von Neumann entropy by

$$S(\rho) = -\text{Tr}(\rho \log_2 \rho) = -\sum_{i=1}^N \lambda_i \log_2 \lambda_i; \quad (2)$$

i.e., it is equal to the Shannon entropy of the eigenvalues of the density matrix, where by convention  $0 \log_2 0 := 0$ .

In this work, we are not limited in having a quantum setup where the matrix  $\rho$  can be called a “density matrix” in the physical sense, but instead we build on this idea to define a matrix from a network that satisfies the same mathematical properties of a density matrix.

First, such a density matrix should be positive semi-definite and symmetric; therefore,  $\rho = Z^{-1} e^{-\beta \mathbf{H}}$ , with  $\mathbf{H}$  a symmetric matrix with non-negative eigenvalues and  $Z$  and  $\beta$  real numbers, is a suitable candidate. Second, the eigenvalues of  $\rho$  must sum to unity, thus imposing the constraint  $Z = \text{Tr} e^{-\beta \mathbf{H}}$ .

If  $\beta$  parametrizes the temperature of a quantum system with Hamiltonian  $\mathbf{H}$  and  $Z = \text{Tr} e^{-\beta \mathbf{H}}$ ,  $\rho$  would provide the Gibbs state of the system in equilibrium at finite temperature. In this scenario, a case corresponding to negative temperatures, more specifically, when  $\beta = -1$ , has been considered. By neglecting the normalizing factor  $Z$  (i.e., by imposing  $Z = 1$ ) and considering  $\mathbf{H} = \mathbf{A}$ , the resulting operator becomes the communicability matrix introduced by Estrada [53–58] (see Ref. [59] for a thorough review).

The concept of communicability has been introduced to provide a quantitative measure of correlation and flow between different parts of a system. Despite the different definitions available in the literature [59], the common rationale is to exploit how information diffuses through a network, beyond the classical shortest-path paradigm. In fact, the communicability matrix is built by considering all possible topological routes between any pairs of nodes in the network and by assigning appropriate (smaller) weights to longer ones.

It has been shown that this approach provides insights about the structure and the function of a system. In fact, communicability can be better understood when modeling a network as a system of classical or quantum oscillators [59] and it is related to the walk entropy used to characterize graphs using statistical-mechanical concepts [60,61].

In this work, we address the problem of defining an information-theoretic framework for complex networks

and, to this aim, we focus on how information diffuses through a network. It is straightforward to show that, if  $\mathbf{p}(0)$  is a vector encoding the amount of information in one or more nodes of the system at time  $\tau = 0$ , its evolution across time is given by  $\mathbf{p}(\tau) = e^{-\tau\mathbf{L}}\mathbf{p}(0)$ , where  $\mathbf{L} = \mathbf{D} - \mathbf{A}$  denotes the combinatorial graph Laplacian, with  $\mathbf{A}$  the adjacency matrix of the network and  $\mathbf{D}$  the diagonal matrix of node degrees. Here,  $e^{-\tau\mathbf{L}}$  is the diffusion propagator at time  $\tau$ .

This propagator exhibits some interesting properties that make it a suitable candidate for a density matrix. Despite the fact that its eigenvalues do not sum to unity, the eigenvalues of the matrix

$$\rho = \frac{e^{-\tau\mathbf{L}}}{Z}, \quad Z = \text{Tr}e^{-\tau\mathbf{L}} \quad (3)$$

satisfy this fundamental property. Therefore, we define the density matrix in terms of the normalized diffusion propagator described above, where  $\beta = \tau > 0$  is interpreted as a parameter playing the role of time. This choice has the advantage of allowing a multiresolution analysis of a system, in the same spirit of recent studies that make use of synchronization dynamics to detect topological scales [62] and Markov processes to unravel the mesoscale organization of complex networks [9,11,12,63–65]. In this framework, it is worth mentioning that  $Z = \sum_{i=1}^N e^{-\lambda_i\tau}$  is proportional to the average return probability, i.e., the probability that a walker starting at any node  $i$  will return to its origin at time  $\tau$ . In the following, we use  $\beta$  and  $\tau$  interchangeably.

In Ref. [66], a matrix is built in such a way that its mathematical properties satisfy the requirements of a density matrix. For an undirected complex network,  $\mathbf{L}$  is symmetric and positive semidefinite, but its eigenvalues do not sum to unity. However, the eigenvalues of the proposed matrix, defined by  $\rho_{\text{BGS}} = \mathbf{L}/\text{Tr}\mathbf{L}$ , evidently do, therefore  $\rho_{\text{BGS}}$  is a mathematically suitable density matrix [38,66] and, hence, Eq. (2) can in principle be applied. This approach has been recently generalized to the case of multilayer systems [52], composite networks where units exhibit different types of relationships that are generally modeled as different layers and used to reduce their structure [47].

However, it has been recently found that the von Neumann entropy calculated from the rescaled Laplacian  $\rho_{\text{BGS}}$  does not satisfy the subadditivity property in some critical circumstances [47]. For instance, let  $\rho_{\text{BGS}}$  be the rescaled Laplacian matrix of a network with  $N$  nodes and  $|E| \gg 1$  edges, let  $\sigma_{\text{BGS}}$  be the rescaled Laplacian matrix of a network with  $N$  nodes and just one (undirected) edge, and let  $\tau_{\text{BGS}}$  be the rescaled Laplacian of the network obtained by summing up, entry wise, the corresponding adjacency matrices of the previous two networks. Since  $S(\sigma_{\text{BGS}}) = 0$ , the subadditivity property  $S(\tau_{\text{BGS}}) \leq S(\rho_{\text{BGS}}) + S(\sigma_{\text{BGS}})$  is not always satisfied because, as we see later,  $S(\tau_{\text{BGS}}) \geq S(\rho_{\text{BGS}})$  very often.

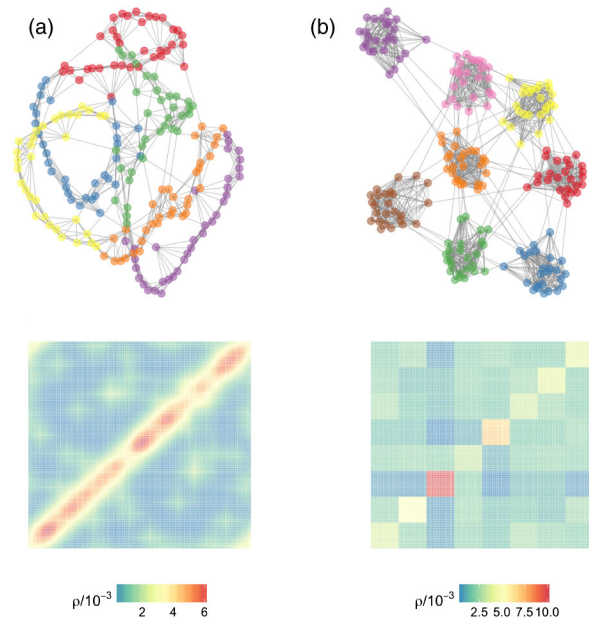


FIG. 1. Density matrix of a complex network. Network (top) and corresponding density matrix (bottom) from (a) a stochastic block model and (b) Watts-Strogatz networks, for  $\beta = 3.16$ . Colors in the networks indicate nodes belonging to the same community, whereas colors in the density matrices indicate the magnitude of the corresponding entry. Networks with  $N = 200$  nodes are shown.

While this peculiar behavior can in fact be exploited in certain situations [47], one is interested in preserving subadditivity, because in general it is expected that the entropy of a composite system  $A + B$  is equal to or smaller than the sum of the entropy of  $A$  and  $B$ . In the case of complex networks, it seems intuitively reasonable but has remained to be proven, see Appendix A, that the aggregation of two graphs is expected to have lower entropy than the sum of the entropy of each graph separately. The entropy of the density matrix introduced to complex networks in this work—which we call spectral entropy—presents some interesting features that we discuss later. A visualization of the density matrix corresponding to different network models is shown in Fig. 1.

### III. VON NEUMANN ENTROPY OF STANDARD NETWORK MODELS

For simplicity, in the following we use  $\rho = \rho_G$ , where there will be no ambiguity about which density matrix is under consideration. Indeed, we also use the notation  $S(G) = S(\rho_G)$  to indicate the entropy of network  $G$ .

From the eigendecomposition of the Laplacian matrix  $\mathbf{L} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1}$ , where  $\mathbf{\Lambda}$  is the diagonal matrix of Laplacian's eigenvalues, it is straightforward to show that

$$Z = \sum_{i=1}^N e^{-\beta\lambda_i(\mathbf{L})}, \quad (4)$$



where  $\lambda_i(\mathbf{L})$  indicates the  $i$ th eigenvalue of  $\mathbf{L}$ . In our framework, the timelike parameter  $\beta$  cannot take negative values. For  $\beta = -1$ , this normalization factor gives the Laplacian Estrada index of a network, which has found applications in graph theory and network science [59,67,68].

The spectral entropy of  $\rho$  can be rewritten as

$$S(\rho) = - \sum_{i=1}^N \lambda_i(\rho) \log_2 \lambda_i(\rho). \quad (5)$$

It follows that

$$\lambda_i(\rho) = Z^{-1} e^{-\beta \lambda_i(\mathbf{L})} \quad (6)$$

provides the relationship between the eigenvalues of the density and the Laplacian matrices. Using Eq. (6), the spectral entropy of the network  $G$  reduces to

$$\begin{aligned} S(G) &= \frac{1}{Z \ln 2} \sum_{i=1}^N e^{-\beta \lambda_i(\mathbf{L})} [\ln Z + \beta \lambda_i(\mathbf{L})] \\ &= \log_2 Z - \beta \frac{\partial \log_2 Z}{\partial \beta} = \log_2 Z + \frac{\beta}{\ln 2} \text{Tr}[\mathbf{L}\rho]. \end{aligned} \quad (7)$$

A similar result, where the negative adjacency matrix  $-\mathbf{A}$  of the graph is used instead of the combinatorial Laplacian, has been obtained from a statistical-mechanical approach to subgraph centrality in complex networks [69]. It is worth remarking that in that context, the parameter  $\beta$  played the role of an inverse temperature, with the operative meaning of rescaling all nodes' interconnections by a factor  $\beta$ .

A possible interpretation of this entropy is given later. Here, it is worth discussing some very special cases, where the spectral entropy following from our definition provides very different results from BGS [ $S_{\text{BGS}} = S(\rho_{\text{BGS}})$ ].

*Networks of isolated nodes.*—First, let us consider the case of a network with no links among nodes. The eigenvalues of the combinatorial Laplacian are all 0,  $Z = N$ , and  $S = \log_2 N$ ; i.e., the entropy is maximum regardless of  $\beta$ , whereas  $S_{\text{BGS}}$  is not defined.

*Single-link networks.*—Let us consider a network with only one link between node  $i$  and node  $j$ ; i.e.,  $\mathbf{A} = \mathbf{E}(ij) + \mathbf{E}(ji)$ , with  $\mathbf{E}(ij)$  the canonical matrix with all zero entries except in  $i$ th- $j$ th component. It is straightforward to show that the eigenvalues of the combinatorial Laplacian are all 0 except the one whose value is 2. While  $S_{\text{BGS}} = 0$ , it follows that

$$S = \log_2 Z + \frac{2\beta e^{-2\beta}}{Z \ln 2}, \quad (8)$$

where  $Z = N - 1 + e^{-2\beta}$  and the asymptotic behavior is  $\log_2 N$  for any  $\beta$ . Our spectral entropy scales with the size of the network when only one directed link is present.

*Lattice graphs.*—Squared lattice graphs with  $N$  nodes, a special case of lattices, can be obtained from the product of two path graphs with  $\sqrt{N}$  nodes. Therefore, the spectrum of the Laplacian matrix is given by the sum of the eigenvalues of the path graphs and are given by  $\lambda_{l,m} = 4\{\sin^2[\pi l/(2\sqrt{N})] + \sin^2[\pi m/(2\sqrt{N})]\}$  ( $l, m = 0, 1, 2, \dots, \sqrt{N} - 1$ ). In this case,

$$Z = \sum_{l,m=1}^{\sqrt{N}} e^{-4\beta[\sin^2(\pi l/2\sqrt{N}) + \sin^2(\pi m/2\sqrt{N})]} \quad (9)$$

and

$$S = \log_2 Z - \frac{\beta}{Z \ln 2} \frac{\partial Z}{\partial \beta}. \quad (10)$$

*Complete networks.*—In the case of the clique network, there are  $N - 1$  eigenvalues equal to  $N$ , from which

$$S = \log_2 Z + \frac{\beta N(Z - 1)}{Z \ln 2}, \quad (11)$$

where  $Z = 1 + (N - 1)e^{-\beta N}$  and the asymptotic behavior is  $\log_2 N$  for small  $\beta$  and 0 for large  $\beta$ .

*Complex network models.*—We show in the top panels of Fig. 2 the von Neumann entropy as a function of  $1/\beta$  for Erdős-Rényi [70], Watts-Strogatz [14], and  $K$ -regular networks, for different values of their parameters, the link probability  $p_{\text{link}}$ , the rewiring probability  $p_{\text{rew}}$ , and the number  $K$  of node neighbors, respectively. In all three network types, when  $\beta$  is high, the entropy tends to zero, whereas it approaches the theoretical limit  $\log_2 N$  when  $\beta$  is small enough. In the bottom panels of Fig. 2, we show the value of the spectral gap as a function of  $1/\beta$ . Intriguingly, the value of  $\beta$ , where the spectral gap is maximum, corresponds to the value where a knee is observed in the spectral entropy (see Appendix B for further details).

*Subadditivity of spectral entropy.*—Let us consider an undirected network  $G$  of  $N$  nodes changing over time, with adjacency matrix  $\mathbf{A}(0)$  at time  $t = 0$ . At each time step  $t$  a pair of two nodes, chosen uniformly randomly and not yet connected, is linked by one undirected link. This is equivalent to having another network  $G'(t)$  consisting of just one link and  $N - 2$  isolated nodes, whose adjacency matrix  $\mathbf{A}'(t)$  is summed up to  $\mathbf{A}(t - 1)$  such that  $\mathbf{A}(t) = \mathbf{A}(t - 1) + \mathbf{A}'(t)$ . If  $G(t - 1)$  and  $G'(t)$  have no edges in common, the above operation is equivalent to the union of the two graphs, another typical approach to aggregate networks.

One immediately asks if the spectral entropy defined in this work and the BGS entropy satisfy the subadditivity property such that

$$S(G(t)) \leq S(G(t - 1)) + S(G'(t)). \quad (12)$$

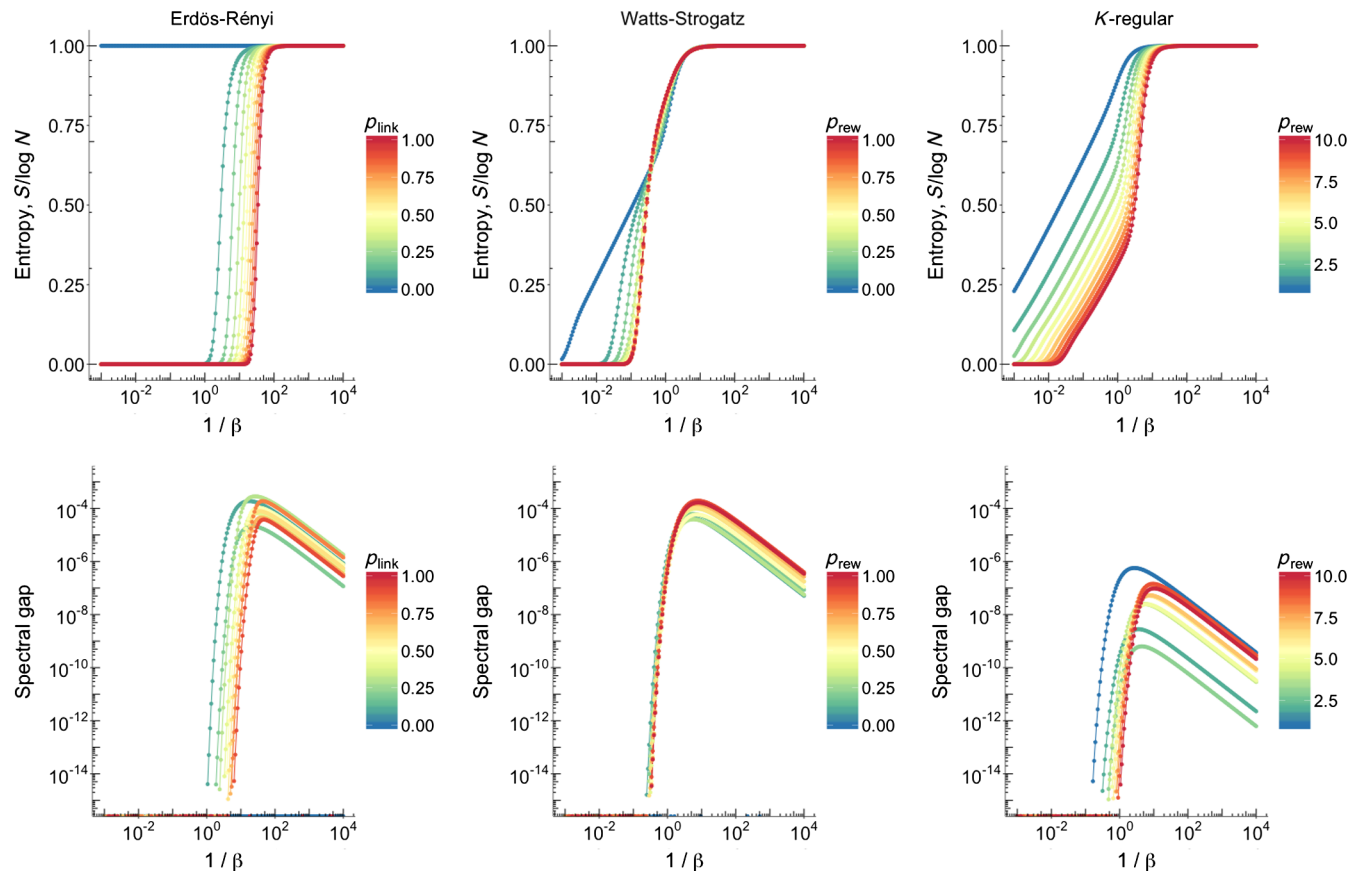


FIG. 2. von Neumann entropy of complex network models. Top panels: Spectral entropy as a function of  $1/\beta$  for Erdős-Rényi networks (left), Watts-Strogatz’s small-world networks (center), and  $K$ -regular lattices (right). Color encodes realizations obtained by varying the parameter of the network model, i.e., the probability  $p_{\text{link}}$  of a link between two nodes, the rewiring probability  $p_{\text{rew}}$ , and the number of neighbors  $K$ , respectively. In all cases, networks with  $N = 200$  are considered. Bottom panels: Spectral gap as a function of  $1/\beta$ . Note that the value of  $\beta$  where it is maximum corresponds to the value where a knee is observed in the spectral entropy (see Appendix B for further details).

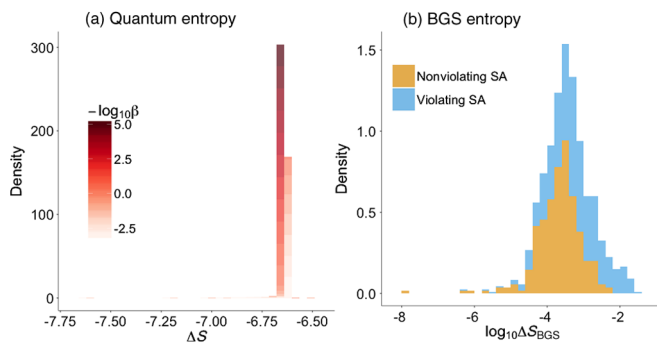


FIG. 3. Subadditivity of spectral entropy. Distribution of entropy variation (see the text) for an Erdős-Rényi network with  $p_{\text{link}} = 0.01$  and  $N = 100$ . The results for the quantum entropy defined in this work (left-hand panel) and the BGS entropy (right-hand panel) are shown. Quantum entropy never violates subadditivity, regardless of the  $\beta$  parameter, whereas the BGS entropy significantly violates subadditivity.

We show in Fig. 3 the distribution of  $\Delta S = S(G(t)) - [S(G(t-1)) + S(G'(t))]$  obtained by calculating both quantum and BGS entropy for an ensemble of Erdős-Rényi networks. Similar results are obtained by using ensembles of Watts-Strogatz and  $K$ -regular networks.

As proven in the Appendix A, the spectral entropy does not violate the subadditivity property regardless of the value of  $\beta$ , whereas the  $S_{\text{BGS}}$  often violates these sensible additivity requirement. See Appendix A for further details.

#### IV. RÉNYI ENTROPY OF COMPLEX NETWORKS

The von Neumann entropy is just a special case of the more general Rényi entropy of a quantum system, defined by [36]

$$S_q = \frac{1}{1-q} \log_2 \text{Tr} \rho^q = \frac{1}{1-q} \log_2 \sum_{i=1}^N \lambda_i(\rho)^q. \quad (13)$$

It is widely used in quantum information theory and quantum computing to quantify entanglement [71] and correlations in physical systems [72], it can be expressed as families of tensor network contractions [73], and it has been found to share a close relationship to free energy [74].

By using Eq. (6), the Rényi spectral entropy of a complex network is therefore given by

$$S_q = \frac{1}{1-q} \ln \sum_{i=1}^N Z^{-q} e^{-q\beta\lambda_i(\mathbf{L})}, \quad (14)$$

in terms of the eigenvalues of the Laplacian matrix. In fact, this entropy generalizes other entropic measures. As  $q$  approaches 0,  $S_q$  increasingly weights all eigenvalues more equally, approaching the Hartley entropy. In the limit of  $q \rightarrow 1$ , Rényi entropy approaches the spectral entropy. As  $q$  approaches  $\infty$ ,  $S_q$  becomes dominated by the high-probability events and it converges to the min entropy. The case with  $q = 2$  recovers the collision entropy—in the case of quantum systems, this is connected to the purity of the system. In fact,  $S_2$  identically vanishes for  $\rho^2 = \rho$ , i.e., for pure states.

In Fig. 4, we show entropy as a function of  $q$  at different values of  $\beta$  for Erdős-Rényi, Watts-Strogatz, and  $K$ -regular networks, respectively, for different values of their parameters. It is common to assess that the Watts-Strogatz network reduces to a  $K$ -regular graph for  $p_{\text{rew}} = 0$  and approaches an Erdős-Rényi network for  $p_{\text{res}} = 1$ . However, the behavior of Rényi entropy as a function of  $q$  shows that there are some significant differences at least in the latter case, where the entropy is considerably larger for the Watts-Strogatz model than for the Erdős-Rényi one.

## V. GENERALIZED QUANTUM DIVERGENCES BETWEEN TWO COMPLEX NETWORKS

One of the main goals of information theory is to quantify the amount of information about a probability distribution—generally obtained from empirical measurements—provided that one has full information about another probability distribution, e.g., the model. This goal is achieved by introducing relative entropies or, equivalently, information divergences.

Similarly, the introduction of divergences (also known as quantum relative entropy) in quantum information theory is foundational to the quest to understand differences between quantum states, quantum and classical information, the quantification of the thermodynamic cost of communication, as well as optimal protocols to transfer information; see, e.g., the reviews [6,75].

The quantum Rényi entropy can be used to define the quantum Rényi divergence, also known as  $q$ -relative Rényi entropy, by

$$\mathcal{D}_q(\rho||\sigma) = \frac{1}{q-1} \log_2 \text{Tr}(\rho^q \sigma^{1-q}), \quad (15)$$

which is defined for  $q > 0$  and reduces to the quantum Kullback-Leibler divergence (or, equivalently, the quantum relative entropy),

$$\mathcal{D}_1(\rho||\sigma) = \text{Tr}[\rho(\log_2 \rho - \log_2 \sigma)], \quad (16)$$

for  $q \rightarrow 1$ .

In general, such divergences are not symmetric and bounded, making difficult certain comparisons. An alternative measure is the quantum  $q$ -Jensen-Shannon divergence [76],

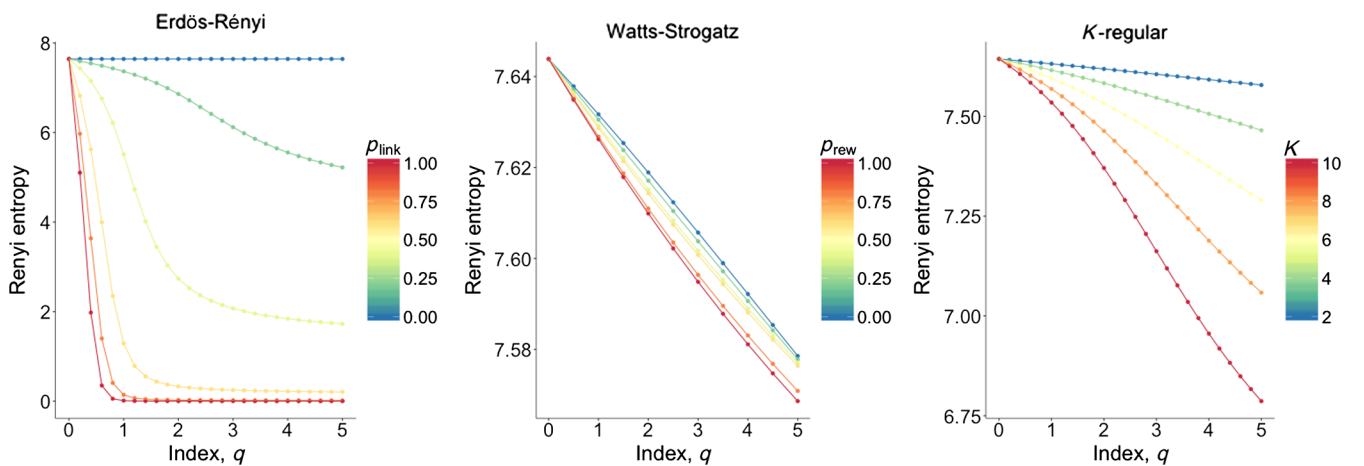


FIG. 4. Rényi entropy of complex network models. Entropy as a function of  $q$  with  $\beta = 1/15.8$  for Erdős-Rényi networks (left), Watts-Strogatz’s small-world networks (center), and  $K$ -regular lattices (right). Color encodes realizations obtained by varying the parameter of the network model, i.e., the probability  $p_{\text{link}}$  of a link between two nodes, the rewiring probability  $p_{\text{rew}}$ , and the number of neighbors  $K$ , respectively. In all cases, networks with  $N = 200$  are considered.

$$\mathcal{J}_q(\boldsymbol{\rho}||\boldsymbol{\sigma}) = S_q\left(\frac{\boldsymbol{\rho} + \boldsymbol{\sigma}}{2}\right) - \frac{1}{2}[S_q(\boldsymbol{\rho}) + S_q(\boldsymbol{\sigma})], \quad (17)$$

where  $\boldsymbol{\mu} = (\boldsymbol{\rho} + \boldsymbol{\sigma}/2)$  is usually called the mixture matrix. It can be proven that  $\mathcal{J}_q^{1/2}$  defines a true metric for  $0 \leq q < 2$  [76–78] and can be used as a measure of distinguishability [47] or similarity [79].

Our formalism allows us to use the family of  $q$ -quantum divergences to compare two networks, and this opens up a new approach to attack two fundamental problems in network science.

### A. Maximum-likelihood estimation and model selection

For a given model and its corresponding set of parameters, the likelihood function measures the probability of observing the data according to the model parameters. Therefore, it is sufficient to perform a maximum-likelihood estimation to obtain the parameters of the model that better reproduce the data, according to the model.

While it is straightforward to define the likelihood function for probability distributions, it is challenging to define a similar concept in the context of density matrices. In the case of complex networks the challenge is complicated by the lack of an appropriate probability distribution. In fact, one usually designs a model, or a set of models, to reproduce only some salient features of an observed complex network, often limiting the comparison between model and data to a few descriptors, such as degree distribution, degree correlations, clustering, or path statistics.

Quantum divergences provide a grounded and unifying approach to this problem. Let us start from the classical problem where an empirical probability distribution  $P(x)$  is obtained from observing the  $N$  outcomes  $\{x_i\}$  ( $i = 1, 2, \dots, N$ ) of a stochastic variable  $\mathcal{X}$ . Let  $Q(x; \Theta)$  be a model to approximate  $P(x)$ , which depends on one (or more) parameter(s), here indicated by  $\Theta$ . In this context, the Kullback-Leibler divergence measures the information gain when the model  $Q(x; \Theta)$  is used to explain the observation  $P(x)$ , and it can be written as

$$\begin{aligned} \mathcal{D}(P||Q) &= \int dx P(x) \log_2 \frac{P(x)}{Q(x; \Theta)} \\ &= -S(P) - \int dx P(x) \log_2 Q(x; \Theta). \end{aligned} \quad (18)$$

If the model  $Q$  is plausible, there exists a value  $\Theta^*$  such that the divergence is minimum, and we are interested in finding such a value by minimizing the divergence with respect to  $\Theta$ . We notice that the first term in the right-hand side of Eq. (18) does not depend on  $\Theta$  and, therefore, plays no role in the minimization procedure. By noticing that

$$P(x) = \frac{1}{N} \sum_{i=1}^N \delta(x - x_i),$$

where  $\delta$  is the Dirac function, the second term in the right-hand side of Eq. (18) reduces to

$$\int dx P(x) \log_2 Q(x; \Theta) = \frac{1}{N} \sum_{i=1}^N \log_2 Q(x_i; \Theta), \quad (19)$$

which is proportional to the negative log-likelihood function. Here, the prefactor can be safely neglected during the minimization procedure. Therefore, by minimizing the Kullback-Leibler divergence, one effectively maximizes the log-likelihood function:

$$\min_{\Theta} \{\mathcal{D}(P||Q)\} = \max_{\Theta} \{\log_2 \mathcal{L}(x; \Theta)\}. \quad (20)$$

We use the proposed framework to achieve a similar result in the case of density matrices. Let  $\boldsymbol{\rho}$  be the density matrix of an empirical network and let  $\boldsymbol{\sigma}(\Theta)$  be a model for such a network, depending on one (or more) parameter(s), here indicated by  $\Theta$ . By starting from Eq. (16), and by arguments similar to the classical case, it is straightforward to show that

$$\min_{\Theta} \{\mathcal{D}(\boldsymbol{\rho}||\boldsymbol{\sigma})\} = \max_{\Theta} \{\text{Tr}[\boldsymbol{\rho} \log_2 \boldsymbol{\sigma}(\Theta)]\}. \quad (21)$$

By comparing the right-hand side of Eqs. (20) and (21), we define the network log-likelihood function by

$$\log_2 \mathcal{L}(\Theta) = \text{Tr}[\boldsymbol{\rho} \log_2 \boldsymbol{\sigma}(\Theta)], \quad (22)$$

where the likelihood function can be calculated by exploiting the properties of the matrix exponential as

$$\begin{aligned} \mathcal{L}(\Theta) &= e^{\text{Tr}[\boldsymbol{\rho} \ln \boldsymbol{\sigma}(\Theta)]} \\ &= \det(e^{\boldsymbol{\rho} \ln \boldsymbol{\sigma}(\Theta)}). \end{aligned} \quad (23)$$

This result allows us to obtain a maximum-likelihood estimation of parameter(s)  $\Theta = \{\theta_1, \theta_2, \dots, \theta_d\}$  by minimizing the Kullback-Leibler divergence between networks. The covariance matrix corresponding to this estimation is the Fisher information matrix (see Appendix C), whose classical counterpart is equivalent to the Hessian of the Kullback-Leibler divergence and it is used to assess the quality of the spectral likelihood estimate. If  $\Theta^*$  is an unbiased estimator of  $\Theta$ , the associated covariance matrix satisfies the Cramer-Rao bound:

$$\text{cov}(\Theta^*) \geq \mathcal{I}^{-1}(\Theta^*). \quad (24)$$

An alternative measure is the Bures-Helstrom metric [80,81], which is proportional to the quantum Fisher



information metric (see Appendix C for the spectral version in the case of complex networks) and can be calculated with a closed expression in terms of eigenvalues and eigenvectors of the density matrix [82].

We can exploit this finding to go beyond model fitting, defining an operative procedure for model selection.

In fact, generally more than one model is used to understand the data, and a fundamental problem in data analysis, known as model selection, is to quantify which model out of a set of candidates is the best one in reproducing the data. One solution to this problem has been given by Akaike, who proposed an information criterion (AIC) for this purpose [83], by showing that the expected value of the relative cross-entropy term in the Kullback-Leibler divergence equals the log-likelihood of the model given the data plus a penalizing constant term that accounts for the number of free parameters. The AIC is given by

$$\text{AIC} = 2k - 2 \log_2 \mathcal{L}(\Theta^*), \quad (25)$$

where  $k$  is the number of parameters of the model and we plug Eq. (22) into Eq. (25) for applications to complex networks. In practice, given a set of models  $\mathcal{M} = \{M_1, M_2, \dots, M_n\}$ , with a number of parameters  $k_1, k_2, \dots, k_n$  and likelihood  $\mathcal{L}_1, \mathcal{L}_2, \dots, \mathcal{L}_n$ , respectively, the most suitable candidate to explain the data is the one being a trade-off between having as small as possible divergence from the data and as small as possible number of parameters, i.e., the one such that AIC is minimum.

Similarly, other model selection criteria can be extended from information theory to the complex network framework. This is the case of Bayesian information criterion (BIC) defined by

$$\text{BIC} = k \log_2 N - 2 \log_2 \mathcal{L}(\Theta^*) \quad (26)$$

and Fisher information approximation (FIA) defined by

$$\text{FIA} = \frac{k}{2} \log_2 \frac{N}{2\pi} - \log_2 \mathcal{L}(\Theta^*) + \log_2 \left( \int d\Theta \sqrt{\det \mathcal{I}(\Theta)} \right), \quad (27)$$

where the last term penalizes a model because of its geometric complexity [84,85], a typical concept in information geometry [86]. In classical statistical inference, it has been shown [85] that the FIA quantifies the length of the shorted description of the data given the model and, as a consequence, its minimization corresponds to finding the minimum description length of the network. The minimum description length principle [87], also known as a formalization of Occam's razor, represents data and models as codes to be compressed, where the model better compresses data to provide its best description. This principle is

one of the most important concepts in information theory and its interpretation, in our context, opens the door for future studies.

To probe the power of the proposed method, we perform several tests on synthetic networks. We generate a network from a specific model with a given value of the parameter(s), assuming it is the observed network, and a set of networks from the same model by varying the value of the parameter(s). Therefore, we perform a maximum-likelihood parameter estimation based on Kullback-Leibler minimization to find the value of the parameter (s) that better fits the observation.

In Fig. 5(a), we consider the case of an Erdős-Rényi network model ( $N = 200$ ), with the link probability  $p_{\text{link}}$  the parameter to fit and  $p_{\text{link}}^* = 0.05$  its expected value. By scanning over the range allowed for  $p_{\text{link}}$ , we sample an ensemble of 100 realizations for each value of the link probability and compare each network against the observed one to calculate the average Kullback-Leibler divergence. In principle, the procedure depends on the value of  $\beta$ ; therefore, we perform the analysis for different values of this parameter in order to understand which value (or range of values) provides the best fit. We consider specific values of  $\beta$  for this purpose; more specifically, we calculate the  $\beta^*$  such that the entropy normalized to its maximum value, i.e.,  $\log_2 N$ , gets a specific real value  $c(\beta^*)$  between 0 and 1. We choose values for  $c(\beta^*)$  ranging between 0.01 and 0.9, and the results show that the global minima correspond to or are very close to the expected value, for  $c(\beta^*) \leq 0.5$ . The best performance is obtained for  $c(\beta^*) < 0.1$ . This analysis suggests a rule of thumb to choose a specific value of  $\beta$  for fitting purposes: the region close to the critical point—where entropy changes from 0 to a positive value provides the most performant range for  $\beta$ .

In Fig. 5(b), we show the Kullback-Leibler divergence of the Watts-Strogatz network model ( $N = 200$ ) for different values of the parameters  $K$  and  $p_{\text{rew}}$  against a Watts-Strogatz network with  $K^* = 6$  and  $p_{\text{rew}}^* = 0.2$ , assumed to be the empirical data. The result shows that the most likely region of the parameter space, i.e., the one where the model is more informative about the data, is successfully identified by the Kullback-Leibler minimization procedure previously described. The result is very interesting because only a single realization of the model, instead of an ensemble as in the previous case, has been used for each pair of parameters, suggesting that the procedure is robust against sample size while reducing the computational cost of the calculation.

Figure 5(c) shows the Kullback-Leibler divergence of a stochastic block model [88] ( $N = 200$ ) for different values of the intra- and intercommunity probability parameters,  $p_{\text{intra}}$  and  $p_{\text{inter}}$ , and number of equally sized blocks against a network obtained from the same model with 8 blocks,  $p_{\text{intra}}^* = 0.5$  and  $p_{\text{inter}}^* = 0.05$ , assumed to be the empirical data. As for the Erdős-Rényi case, we sample an ensemble



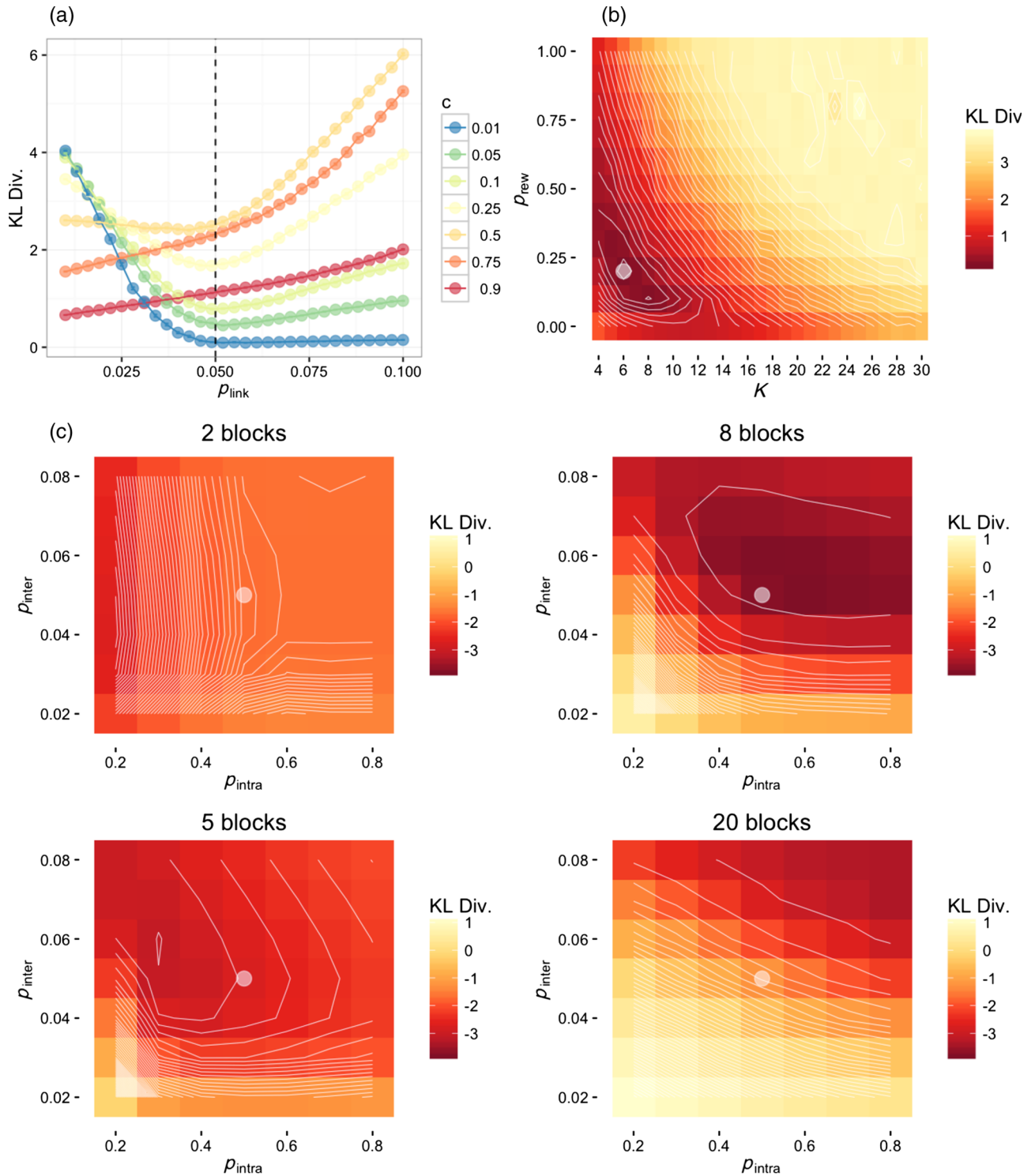


FIG. 5. Maximum-likelihood parameter estimation based on Kullback-Leibler (KL) minimization. (a) Erdős-Rényi network model ( $N = 200$ ) for different values of the link probability  $p_{\text{link}}$  against an Erdős-Rényi network with  $p_{\text{link}}^* = 0.05$ , assumed to be the data to fit. The vertical dashed line indicates the true value of the parameter. Each curve corresponds to a value  $\beta^*$  of  $\beta$  such that  $S[\rho(\beta^*)]/\log_2 N = c(\beta^*)$ , with  $c$  a real number between 0 and 1. Global minima are expected around the true value. (b) Watts-Strogatz's network model ( $N = 200$ ) for different values of the parameters  $K$  and  $p_{\text{rew}}$  against a Watts-Strogatz network with  $K^* = 6$  and  $p_{\text{rew}}^* = 0.2$ , assumed to be the data to fit. The white dot indicates the true value and it falls in the iso-likelihood region with smallest Kullback-Leibler divergence (encoded by color in log scale). (c) Stochastic block model ( $N = 200$ ) for different values of the intra- and intercommunity probability parameters,  $p_{\text{intra}}$  and  $p_{\text{inter}}$ , and number of equally sized blocks against a network obtained from the same model with 8 blocks,  $p_{\text{intra}}^* = 0.5$  and  $p_{\text{inter}}^* = 0.05$ , assumed to be the data to fit.

of 100 realizations for each triad of parameters to calculate the average Kullback-Leibler divergence between model and an observation. We show the results for different block sizes, and the overall minimum is found for 8 blocks,  $p_{\text{intra}} = 0.6$  and  $p_{\text{inter}} = 0.05$ , in excellent agreement with expectation.

### B. Clustering layers of multilayer systems

The second application concerns the more recent problem of identifying layers of a multiplex network or snapshots of a time-varying network [52] that provide redundant information about the system and that might be aggregated to reduce the complexity of the system [47]. In this case, by using appropriate quantum divergences, it is possible to calculate a distance between layers or snapshots and devise exact or heuristics procedures to cluster them. With the recent rise of interest in multilayer systems [49–51], this problem became very relevant for practical applications, e.g., to demonstrate the validity of a multilayer approach to modeling complex systems such as the human brain [89] or to reveal the interplay between collaboration networks and the organization of knowledge in physics [90], with new approaches to cluster, aggregate, or interconnect layers proposed more recently [90–92].

Here, we use the networks built from analysis of the structure and function of the human microbiome, consisting of 18 layers of a multiplex network, each one corresponding to a body site. Recently, such layers have been partitioned into community types, by using Dirichlet multinomial mixture models, that may be associated with complex diseases [53]. We use the same data and calculate the Jensen-Shannon distance between each pair of layers for different values of the  $\beta$  parameter. We show in Fig. 6 the

resulting hierarchical clustering of the layers, in good agreement with the result reported in Ref. [53]. A more quantitative comparison against the results in Ref. [53] is difficult because only  $p$ -value upper bounds are reported in the study for community-type associations. Nevertheless, our method is able to correctly reproduce the clustering of sites within certain body regions: for instance, the community consisting of hard palate, tongue dorsum, saliva, palatine tonsils, and throat, as well as the community made by vaginal introitus, mid vagina, and posterior fornix, and the smaller ones made by sub- and supragingival plaque and left or right retroauricular crease. For ranges of  $\beta$  within 1 order of magnitude, the differences between hierarchies are small and not significant. Here, we show the results for  $\beta = 0.1$  and  $\beta = 10$ , with cophenetic distance equal to 0.84 and Baker’s gamma correlation coefficient equal to 0.68. By performing the same analysis on hierarchies where labels are reshuffled uniformly random, we reject the null hypothesis that the two dendrograms are uncorrelated ( $P < 0.001$ ), supporting the fact that the result is robust to the choice of  $\beta$ .

### VI. DISCUSSION

The use of von Neumann entropy is central to modern quantum information theory, with many new insights, uses, and interpretations arising often. In the context of complex networks, spectral entropy might be considered as a measure of the information content of the system, although this interpretation is not easy to accept without the appropriate formalization typical of classical and quantum information theory.

Here, we observe that networks of isolated nodes (which challenge the notion of a network itself) and fully connected networks, i.e., cliques, have maximal entropy (the

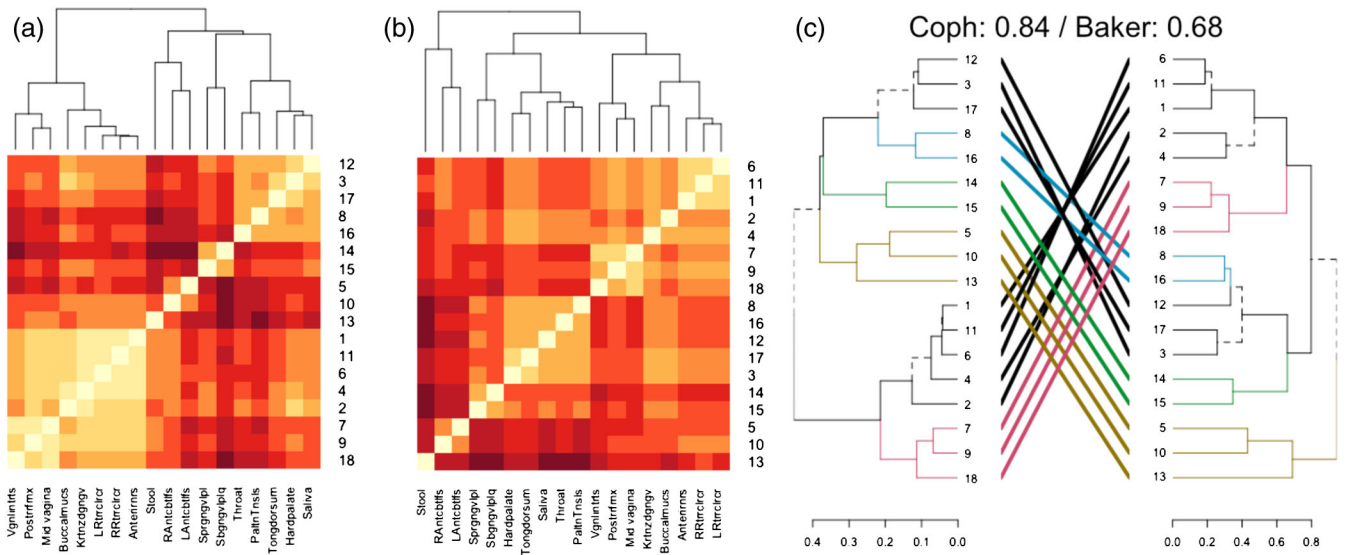


FIG. 6. Hierarchical clustering of human microbiome sites. The Jensen-Shannon distance matrices with (a)  $\beta = 0.1$  and (b)  $\beta = 10$  are shown, together with their (c) tanglegram, i.e., a visual analysis of their differences. Cophenetic distance and Baker’s gamma correlation coefficient, quantifying the correlation between the two dendrograms, are reported.

latter when considered at small values of diffusion time). When diffusion time is large enough, the entropy of cliques goes to 0, suggesting that fully connected systems are the most ordered (or least uncertain). We investigate several alternative explanations for the entropy proposed in this work, but the most promising one is based on the number of possible configurations obtained by reshuffling the connections among nodes, which is equivalent to randomly reshuffling the entries of the adjacency matrix. In fact, all possible reshuffled configurations of both the network of isolated nodes and the clique network do not alter the system. In other words, the graph isomorphism problem becomes trivial, as all possible networks are not distinguishable from each other, thus maximizing the uncertainty about the system and, consequently, the entropy. In the case of  $K$ -regular networks, the situation is a bit different, because not all reshuffled configurations will keep them unaltered; therefore, such networks are expected to exhibit lower entropy than  $\log_2 N$ . The case of an Erdős-Rényi network, which is characterized by stochastically homogeneous connectivity, is interesting. The entropy for this type of network is almost always zero or  $\log_2 N$ , except for a narrow range of  $\beta$ . This is compatible with our interpretation, because on average almost any reshuffled configuration of this type of network will resemble the original. According to this interpretation, extremely sparse networks and extremely dense networks should exhibit almost maximum entropy (the latter for insufficient diffusion time). We perform additional numerical experiments and theoretical analysis of toy models (path graphs, lattices, etc.) and we verify our expectation.

The proposed framework is entirely based on the computation of eigenvalues of density matrices and their products. The computational complexity of such an operation is generally expected to be polynomial and to scale as  $\mathcal{O}(n^\gamma)$ , with  $2 < \gamma < 2.4$ , making challenging calculations for networks with hundreds of thousands of nodes. However, recent advances in parallel operations on dense symmetric matrices make it possible to solve eigenvalue problems for matrices with size of the order of  $10^5$  in less than 5 min on standard computers [93].

## VII. CONCLUSIONS AND OUTLOOK

The concept of thermodynamic entropy has been crucial to understanding the structure and the dynamics of complex systems. The generalization of this concept to quantum mechanics by von Neumann was a milestone in the field, allowing us, among other applications, to characterize the mixedness of quantum states.

An appropriate definition of entropy has been lacking in the field of complex networks, mainly because there is no simple way to define a probability distribution able to represent, without loss of information, the network. Previous attempts to define the entropy of a complex network were mainly based on the calculation of

Shannon entropy of the probability distribution of some descriptor(s), while neglecting other information.

More recently, the idea that a quantumlike entropy might be introduced for complex networks has been explored. Here, we demonstrate that previous attempts fail in preserving fundamental properties of entropy, like subadditivity. Motivated by this finding, we propose a density matrix closely related to how information diffuses through a network, which nontrivially depends on the corresponding combinatorial Laplacian matrix. One of the main advantages of this approach, rooted in spectral theory, is that the resulting entropy does not depend on the distribution of a specific network descriptor, but it depends on the network as a whole.

In this study, we show, analytically and numerically, that our density matrix allows us to preserve the basic properties of an entropy. More specifically, we demonstrate that the entropy of the system obtained by aggregating two different networks must be equal to or smaller than the sum of the entropy of the two nonaggregated networks.

This definition of entropy allows us to define Rényi spectral entropy and generalize relative entropies, also known as quantum divergences. By using Kullback-Leibler divergence and well-known results in classical information theory, we devise a maximum-likelihood approach to model fitting, entirely based on the eigenvalue spectra of density matrices. Our numerical experiments confirm that by minimizing the Kullback-Leibler divergence between observed network and models, we obtain the maximum spectral likelihood estimation of parameters. This result might be of special interest for applications devoted to unraveling the mesoscale organization of a network, among others.

Finally, by using the Jensen-Shannon divergence, whose square root provides a metric, we show that our framework can be applied to quantify the spectral distance between two networks. More specifically, this result is of interest for applications in multilayer network research, where the distance among layers can be measured and used to hierarchically cluster them. As a representative example of the power of our methodology, we show that this approach recovers with excellent accuracy the community-based classification of human microbiome sites.

This study opens interesting perspectives on future works. For instance, while our inference procedure does not provide the community label to assign to each node, as in community-detection methods based on stochastic block modeling [94–100], it provides a fast way, based on a network's spectral properties, to explore the parameter space in order to identify the most informative region about the data. An intriguing possibility to explore is how this result can complement existing network inference approaches [101].

A future application of our framework—of great interest for the community of network scientists—is the quantification of how much information is needed for correctly

learning the parameters of a model, which should not be confused with the Cramer-Rao bound. This represents a crucial issue in network science: connectivity of empirical networks is often sampled through *ad hoc* algorithms, as in the case of virtual social systems, the internet or the world wide web. Quantifying to what extent it is possible to learn the parameter(s) of a network model in the absence of partial connectivity information is a tantalizing possibility, expected to deepen our understanding of networked systems.

The information-theoretic quantities we introduce in this work open the intriguing possibility to extend information geometry to complex networks. In information geometry parametric statistical models define a Riemannian manifold endowed with a Fisher information matrix as a metric tensor [102,103]. The importance of information manifolds for statistical inference is well established and has found applications to machine learning [86,104,105] and phase transitions in quantum systems [106–111]. It will be interesting to explore to what extent the success of information geometry can be ported to statistical inference of complex networks, where the metric tensor of the underlying information manifold is given by the spectral Fisher information matrix.

The potential of the proposed methodologies goes beyond network science and we envision important contributions to biology—where it has been shown that network entropy can be used to identify lethal proteins [112], reveal cancer system hallmarks [113], and characterize aging and cancer progression [114]—and to quantum physics—especially to the emergent field of Hamiltonian learning [115,116] and inference of quantum complex network models based on qubit entanglement [29,30].

## ACKNOWLEDGMENTS

The authors thanks Alex Arenas, Mason A. Porter, and Mario Chavez for useful discussions. M. D. D. acknowledges financial support from the Spanish program Juan de la Cierva (IJCI-2014-20225). J. B. acknowledges IARPA and the Foundational Questions Institute (FQXi) for financial support.

## APPENDIX A: SUBADDITIVITY OF THE VON NEUMANN ENTROPY FOR AGGREGATE NETWORKS

Let  $\mathbf{A}$  and  $\mathbf{B}$  be the adjacency matrices of two networks  $G_A$  and  $G_B$ , respectively, with  $N$  nodes. Let  $\mathbf{C} = \mathbf{A} + \mathbf{B}$  be the adjacency matrix obtained by their sum, which represents a new network  $G_C$  obtained by aggregating  $G_A$  and  $G_B$ . In the following, we show that  $S(G_C) \leq S(G_A) + S(G_B)$ , i.e., that the subadditivity property of the entropy is always satisfied. For the sake of simplicity, we use the notation  $S_X = S(G_X)$ .

Let us indicate with  $\mathbf{L}_A$ ,  $\mathbf{L}_B$ , and  $\mathbf{L}_C$  the Laplacian matrices of  $G_A$ ,  $G_B$ , and  $G_C$ , respectively, and let  $\rho_A$ ,  $\rho_B$ , and  $\rho_C$  indicate the corresponding density matrices defined as in Eq. (3). In the following, we work in nat units, instead of bits.

The Kullback-Leibler divergence of two density matrices is always non-negative, as a consequence of Klein's inequality with  $f(\mathbf{X}) = \mathbf{X} \ln \mathbf{X}$ . Therefore,

$$\begin{aligned} \mathcal{D}(\rho_C || \rho_A) &= -S_C - \text{Tr}[\rho_C \ln \rho_A] \\ &= -S_C + \beta \text{Tr}[\mathbf{L}_A \rho_C] + \ln Z_A \geq 0, \end{aligned} \quad (\text{A1})$$

and similarly for  $\mathcal{D}(\rho_C || \rho_B)$ . As  $\mathbf{L}_A$ ,  $\mathbf{L}_B$ ,  $\rho_A$ , and  $\rho_B$  are positive semidefinite, from their Cholesky factorization it is straightforward to show that

$$\begin{aligned} \text{Tr}[\mathbf{L}_X \rho_X] &= \text{Tr}[(\mathbf{D}\mathbf{D}^\dagger)(\mathbf{Q}\mathbf{Q}^\dagger)] \\ &= \text{Tr}[(\mathbf{Q}^\dagger \mathbf{D})(\mathbf{Q}^\dagger \mathbf{D})^\dagger] \geq 0. \end{aligned} \quad (\text{A2})$$

Moreover, it is possible to show that  $\ln Z_X \geq 0$ . In fact,  $Z_X = \sum_{i=1}^N e^{-\beta \lambda_i(\mathbf{L}_X)}$ , and  $\lambda_1(\mathbf{L}_X) = 0$ , because of the Perron-Frobenius theorem. Therefore,

$$Z_X = 1 + \sum_{i=2}^N e^{-\beta \lambda_i(\mathbf{L}_X)} \geq 1, \quad (\text{A3})$$

from which  $\ln Z_X \geq 0$ .

By summing up such non-negative terms, the following inequality,

$$\begin{aligned} \mathcal{D}(\rho_C || \rho_A) + \mathcal{D}(\rho_C || \rho_B) \\ + \beta \text{Tr}[\mathbf{L}_A \rho_A] + \beta \text{Tr}[\mathbf{L}_B \rho_B] + \ln Z_C \geq 0, \end{aligned} \quad (\text{A4})$$

holds. The above inequality can be expanded into

$$\begin{aligned} -S_C + \beta \text{Tr}[\mathbf{L}_A \rho_C] + \ln Z_A \\ + -S_C + \beta \text{Tr}[\mathbf{L}_B \rho_C] + \ln Z_B \\ + \beta \text{Tr}[\mathbf{L}_A \rho_A] + \beta \text{Tr}[\mathbf{L}_B \rho_B] + \ln Z_C \geq 0. \end{aligned} \quad (\text{A5})$$

By exploiting the fact that the von Neumann entropy is given by

$$S(\rho_X) = \beta \text{Tr}[\mathbf{L}_X \rho_X] + \ln Z_X \quad (\text{A6})$$

and  $\mathbf{L}_C = \mathbf{L}_A + \mathbf{L}_B$ , it follows that

$$S_A + S_B - 2S_C + \ln Z_C + \beta \text{Tr}[\mathbf{L}_C \rho_C] \geq 0,$$

which leads to  $S_A + S_B \geq S_C$ .



## APPENDIX B: RELATIONSHIP BETWEEN SPECTRAL ENTROPY AND SPECTRAL GAP

From the spectral decomposition of the density matrix, it is straightforward to show that the spectral entropy of a network can be written as [see Eq. (7)]

$$S(G) = \log_2 \sum_{i=1}^N e^{-\beta \lambda_i(\mathbf{L})} + \frac{\beta}{\ln 2} \frac{\sum_{i=1}^N \lambda_i(\mathbf{L}) e^{-\beta \lambda_i(\mathbf{L})}}{\sum_{i=1}^N e^{-\beta \lambda_i(\mathbf{L})}}, \quad (\text{B1})$$

where  $\lambda_i(\mathbf{L})$  is the  $i$ th eigenvalue of the Laplacian matrix of the network  $G$ . The spectral gap of the density matrix  $\rho$  is given by

$$\Delta_\beta = \lambda_2(\rho) - \lambda_1(\rho), \quad (\text{B2})$$

and the value  $\beta^*$ , where it is maximum, can be calculated (when it exists) by imposing the first derivative equal to zero and by verifying that the second derivative in the same point is smaller than zero. It is straightforward to show that at  $\beta^*$  the first derivative leads to

$$S^* = \log_2 \sum_{i=1}^N e^{-\beta^* \lambda_i(\mathbf{L})} - \frac{\beta^*}{\ln 2} \frac{\lambda_N(\mathbf{L}) e^{-\beta^* \lambda_N(\mathbf{L})} - \lambda_{N-1}(\mathbf{L}) e^{-\beta^* \lambda_{N-1}(\mathbf{L})}}{e^{-\beta^* \lambda_N(\mathbf{L})} - e^{-\beta^* \lambda_{N-1}(\mathbf{L})}},$$

where  $S^*$  is the value of the spectral entropy at  $\beta^*$ , where the spectral gap is maximum. This maximum is not guaranteed to exist, but in all the examples we probe numerically, it did. By indicating with  $Z^*$  the normalizing factor for  $\beta = \beta^*$ , the above equation can be reduced to

$$S^* = \log_2 Z^* + \frac{\beta^*}{\Delta_{\beta^*} \ln 2} [\lambda_N(\mathbf{L}) \lambda_1^*(\rho) - \lambda_{N-1}(\mathbf{L}) \lambda_2^*(\rho)].$$

For  $\beta < \beta^*$ , the second term on the rhs rapidly approaches zero, whereas  $\log_2 Z$  approaches  $\log_2 N$ , i.e., the maximum attainable entropy. Therefore, the entropy is maximum (or close enough to its maximum value) before the spectral gap (before and if) reaches its maximum, as a function of  $\beta$ .

## APPENDIX C: SPECTRAL FISHER INFORMATION MATRIX

Let  $\rho(\Theta)$  be the parametric statistical model, depending on the parameter set  $\Theta = \{\theta_1, \theta_2, \dots, \theta_d\}$ , for a complex network with density matrix  $\rho$ . In quantum physics, the upper bound to the expected Fisher information is given by the quantum information [81,106,117],

$$\mathcal{I}_{\alpha\beta}(\Theta) = \mathbb{E}(\Sigma^{(\alpha)} \circ \Sigma^{(\beta)}) = \text{Tr}[\rho \Sigma^{(\alpha)} \circ \Sigma^{(\beta)}], \quad (\text{C1})$$

where  $\Sigma^{(\alpha)}$  is the symmetric logarithmic derivative—with respect to the  $\alpha$ th parameter ( $\alpha = 1, 2, \dots, d$ )—defined by

$$\frac{\partial}{\partial \theta_\alpha} \rho(\Theta) = \Sigma^{(\alpha)} \circ \rho(\Theta), \quad (\text{C2})$$

and we use the symmetric product defined by

$$\mathbf{X} \circ \mathbf{Y} = \frac{1}{2} (\mathbf{X}\mathbf{Y} + \mathbf{Y}\mathbf{X}). \quad (\text{C3})$$

Let us consider the spectral decomposition of the density matrix

$$\rho(\Theta) = Z(\Theta)^{-1} \mathbf{Q}(\Theta) e^{-\beta \Lambda(\Theta)} \mathbf{Q}^{-1}(\Theta), \quad (\text{C4})$$

where  $\Lambda(\Theta)$  is the diagonal matrix of Laplacian's eigenvalues  $\lambda_i(\Theta)$  ( $i = 1, 2, \dots, N$ ) and the columns of the matrix  $\mathbf{Q}(\Theta)$  are the corresponding eigenvectors, which we indicate by  $\mathbf{q}_i(\Theta)$ . Note that we do not indicate explicitly the dependence on  $\mathbf{L}$ , for the sake of simplicity. In the Laplacian eigenbasis, Eq. (C2) reads

$$\mathbf{q}_i^\top(\Theta) \frac{\partial}{\partial \theta_\alpha} \rho(\Theta) \mathbf{q}_j(\Theta) = \frac{1}{2} [\lambda_i(\Theta) + \lambda_j(\Theta)] \Sigma_{ij}^{(\alpha)},$$

from which we obtain the entries of the symmetric logarithmic derivative [118] as

$$\Sigma_{ij}^{(\alpha)} = \frac{2\Phi_{ij}^{(\alpha)}(\Theta)}{\lambda_i(\Theta) + \lambda_j(\Theta)}, \quad (\text{C5})$$

where

$$\Phi_{ij}^{(\alpha)}(\Theta) = \frac{\partial}{\partial \theta_\alpha} \lambda_i(\Theta) \delta_{ij} + (\lambda_j - \lambda_i) \mathbf{q}_i^\top(\Theta) \frac{\partial}{\partial \theta_\alpha} \mathbf{q}_j(\Theta),$$

and  $\delta_{ij}$  is the Kronecker function.

- 
- [1] C. Shannon, *A Mathematical Theory of Communication*, *Bell Syst. Tech. J.* **27**, 379 (1948).
  - [2] A. Borst and F. E. Theunissen, *Information Theory and Neural Coding*, *Nat. Neurosci.* **2**, 947 (1999).
  - [3] S. P. Strong, R. Koberle, R. R. d. R. van Steveninck, and W. Bialek, *Entropy and Information in Neural Spike Trains*, *Phys. Rev. Lett.* **80**, 197 (1998).
  - [4] C. H. Bennett and P. W. Shor, *Quantum Information Theory*, *IEEE Trans. Inf. Theory* **44**, 2724 (1998).
  - [5] A. S. Holevo, *Bounds for the Quantity of Information Transmitted by a Quantum Communication Channel*, *Prob. Peredachi Inf.* **9**, 3 (1973).
  - [6] V. Vedral, *The Role of Relative Entropy in Quantum Information Theory*, *Rev. Mod. Phys.* **74**, 197 (2002).
  - [7] P. Bernaola-Galván, I. Grosse, P. Carpena, J. L. Oliver, R. Román-Roldán, and H. E. Stanley, *Finding Borders between Coding and Noncoding DNA Regions by an Entropic Segmentation Method*, *Phys. Rev. Lett.* **85**, 1342 (2000).

- [8] R. N. Mantegna, S. V. Buldyrev, A. L. Goldberger, S. Havlin, C.-K. Peng, M. Simons, and H. E. Stanley, *Linguistic Features of Noncoding DNA Sequences*, *Phys. Rev. Lett.* **73**, 3169 (1994).
- [9] M. De Domenico, A. Lancichinetti, A. Arenas, and M. Rosvall, *Identifying Modular Flows on Multilayer Networks Reveals Highly Overlapping Organization in Social Systems*, *Phys. Rev. X* **5**, 011027 (2015).
- [10] A. V. Esquivel and M. Rosvall, *Compression of Flow Can Reveal Overlapping-Module Organization in Networks*, *Phys. Rev. X* **1**, 021025 (2011).
- [11] M. Rosvall and C. T. Bergstrom, *Maps of Random Walks on Complex Networks Reveal Community Structure*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1118 (2008).
- [12] M. Rosvall, A. V. Esquivel, A. Lancichinetti, J. D. West, and R. Lambiotte, *Memory in Network Flows and Its Effects on Spreading Dynamics and Community Detection*, *Nat. Commun.* **5**, 4630 (2014).
- [13] A.-L. Barabási and R. Albert, *Emergence of Scaling in Random Networks*, *Science* **286**, 509 (1999).
- [14] D. J. Watts and S. H. Strogatz, *Collective Dynamics of 'Small-World' Networks*, *Nature (London)* **393**, 440 (1998).
- [15] M. Boguna, D. Krioukov, and K. C. Claff, *Navigability of Complex Networks*, *Nat. Phys.* **5**, 74 (2009).
- [16] V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani, *Detecting Rich-Club Ordering in Complex Networks*, *Nat. Phys.* **2**, 110 (2006).
- [17] R. Guimera and L. A. N. Amaral, *Functional Cartography of Complex Metabolic Networks*, *Nature (London)* **433**, 895 (2005).
- [18] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, *Uncovering the Overlapping Community Structure of Complex Networks in Nature and Society*, *Nature (London)* **435**, 814 (2005).
- [19] C. Song, S. Havlin, and H. A. Makse, *Self-Similarity of Complex Networks*, *Nature (London)* **433**, 392 (2005).
- [20] A. Vespignani, *Modelling Dynamical Processes in Complex Socio-Technical Systems*, *Nat. Phys.* **8**, 32 (2012).
- [21] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, *Controllability of Complex Networks*, *Nature (London)* **473**, 167 (2011).
- [22] R. Albert, H. Jeong, and A.-L. Barabási, *Error and Attack Tolerance of Complex Networks*, *Nature (London)* **406**, 378 (2000).
- [23] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, *Catastrophic Cascade of Failures in Interdependent Networks*, *Nature (London)* **464**, 1025 (2010).
- [24] D. S. Callaway, M. E. Newman, S. H. Strogatz, and D. J. Watts, *Network Robustness and Fragility: Percolation on Random Graphs*, *Phys. Rev. Lett.* **85**, 5468 (2000).
- [25] M. De Domenico, A. Solé-Ribalta, S. Gómez, and A. Arenas, *Navigability of Interconnected Networks under Random Failures*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8351 (2014).
- [26] J. Gao, S. V. Buldyrev, H. E. Stanley, and S. Havlin, *Networks Formed from Interdependent Networks*, *Nat. Phys.* **8**, 40 (2012).
- [27] F. Radicchi and A. Arenas, *Abrupt Transition in the Structural Formation of Interconnected Networks*, *Nat. Phys.* **9**, 717 (2013).
- [28] A. Acín, J. I. Cirac, and M. Lewenstein, *Entanglement Percolation in Quantum Networks*, *Nat. Phys.* **3**, 256 (2007).
- [29] M. Cuquet and J. Calsamiglia, *Entanglement Percolation in Quantum Complex Networks*, *Phys. Rev. Lett.* **103**, 240503 (2009).
- [30] S. Perseguers, M. Lewenstein, A. Acín, and J. I. Cirac, *Quantum Random Networks*, *Nat. Phys.* **6**, 539 (2010).
- [31] M. Faccin, T. Johnson, J. Biamonte, S. Kais, and P. Migdał, *Degree Distribution in Quantum Walks on Complex Networks*, *Phys. Rev. X* **3**, 041007 (2013).
- [32] M. Faccin, P. Migdał, T. H. Johnson, V. Bergholm, and J. D. Biamonte, *Community Detection in Quantum Complex Networks*, *Phys. Rev. X* **4**, 041012 (2014).
- [33] D. Deutsch, *Quantum Theory, the Church-Turing Principle and the Universal Quantum Computer*, *Proc. R. Soc. A* **400**, 97 (1985).
- [34] H. Araki and E. H. Lieb, *Entropy Inequalities*, *Commun. Math. Phys.* **18**, 160 (1970).
- [35] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, *Quantum Entanglement*, *Rev. Mod. Phys.* **81**, 865 (2009).
- [36] A. Wehrl, *General Properties of Entropy*, *Rev. Mod. Phys.* **50**, 221 (1978).
- [37] M. M. Wilde, *Quantum Information Theory* (Cambridge University Press, Cambridge, 2013).
- [38] K. Anand, G. Bianconi, and S. Severini, *Shannon and von Neumann Entropy of Random Networks with Heterogeneous Expected Degree*, *Phys. Rev. E* **83**, 036109 (2011).
- [39] G. Bianconi, P. Pin, and M. Marsili, *Assessing the Relevance of Node Features for Network Structure*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 11433 (2009).
- [40] M. Dehmer, *Information-Theoretic Concepts for the Analysis of Complex Networks*, *Appl. Artif. Intell.* **22**, 684 (2008).
- [41] N. Eagle, M. Macy, and R. Claxton, *Network Diversity and Economic Development*, *Science* **328**, 1029 (2010).
- [42] S. Johnson, J. J. Torres, J. Marro, and M. A. Munoz, *Entropic Origin of Disassortativity in Complex Networks*, *Phys. Rev. Lett.* **104**, 108702 (2010).
- [43] I. A. Kovács, R. Mizsei, and P. Csermely, *A Unified Data Representation Theory for Network Visualization, Ordering and Coarse-Graining*, *Sci. Rep.* **5**, 13786 (2015).
- [44] N. Lüdtke, S. Panzeri, M. Brown, D. S. Broomhead, J. Knowles, M. A. Montemurro, and D. B. Kell, *Information-Theoretic Sensitivity Analysis: A General Method for Credit Assignment in Complex Networks*, *J. R. Soc. Interface* **5**, 223 (2008).
- [45] M. Rosvall and C. T. Bergstrom, *An Information-Theoretic Framework for Resolving Community Structure in Complex Networks*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7327 (2007).
- [46] M. Rosvall and C. T. Bergstrom, *Mapping Change in Large Networks*, *PLoS One* **5**, e8694 (2010).

- [47] M. De Domenico, V. Nicosia, A. Arenas, and V. Latora, *Structural Reducibility of Multilayer Networks*, *Nat. Commun.* **6**, 6864 (2015).
- [48] J.-P. Onnela, D. J. Fenn, S. Reid, M. A. Porter, P. J. Mucha, M. D. Fricker, and N. S. Jones, *Taxonomies of Networks from Community Structure*, *Phys. Rev. E* **86**, 036104 (2012).
- [49] S. Boccaletti, G. Bianconi, R. Criado, C. Del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, and M. Zanin, *The Structure and Dynamics of Multilayer Networks*, *Phys. Rep.* **544**, 1 (2014).
- [50] M. De Domenico, C. Granell, M. A. Porter, and A. Arenas, *The Physics of Spreading Processes in Multilayer Networks*, *Nat. Phys.* **12**, 901 (2016).
- [51] M. Kivela, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, *Multilayer Networks*, *J. Complex Netw.* **2**, 203 (2014).
- [52] M. De Domenico, A. Solè-Ribalta, E. Cozzo, M. Kivela, Y. Moreno, M. A. Porter, S. Gómez, and A. Arenas, *Mathematical Formulation of Multilayer Networks*, *Phys. Rev. X* **3**, 041022 (2013).
- [53] T. Ding and P. D. Schloss, *Dynamics and Associations of Microbial Community Types across the Human Body*, *Nature (London)* **509**, 357 (2014).
- [54] E. Estrada and J. Gómez-Gardeñes, *Communicability Reveals a Transition to Coordinated Behavior in Multiplex Networks*, *Phys. Rev. E* **89**, 042819 (2014).
- [55] E. Estrada and N. Hatano, *Communicability in Complex Networks*, *Phys. Rev. E* **77**, 036111 (2008).
- [56] E. Estrada and D. J. Higham, *Network Properties Revealed through Matrix Functions*, *SIAM Rev.* **52**, 696 (2010).
- [57] E. Estrada, D. J. Higham, and N. Hatano, *Communicability and Multipartite Structures in Complex Networks at Negative Absolute Temperatures*, *Phys. Rev. E* **78**, 026102 (2008).
- [58] P. Grindrod, M. C. Parsons, D. J. Higham, and E. Estrada, *Communicability across Evolving Networks*, *Phys. Rev. E* **83**, 046120 (2011).
- [59] E. Estrada, N. Hatano, and M. Benzi, *The Physics of Communicability in Complex Networks*, *Phys. Rep.* **514**, 89 (2012).
- [60] E. Estrada and A. José, *Maximum Walk Entropy Implies Walk Regularity*, *Linear Algebra Appl.* **458**, 542 (2014).
- [61] E. Estrada, A. José, and N. Hatano, *Walk Entropies in Graphs*, *Linear Algebra Appl.* **443**, 235 (2014).
- [62] A. Arenas, A. Díaz-Guilera, and C. J. Pérez-Vicente, *Synchronization Reveals Topological Scales in Complex Networks*, *Phys. Rev. Lett.* **96**, 114102 (2006).
- [63] J. Delvenne, S. Yaliraki, and M. Barahona, *Stability of Graph Communities across Time Scales*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 12755 (2010).
- [64] R. Lambiotte, J.-C. Delvenne, and M. Barahona, *Random Walks, Markov Processes and the Multiscale Modular Organization of Complex Networks*, *IEEE Trans. Netw. Sci. Eng.* **1**, 76 (2014).
- [65] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, *Community Structure in Time-Dependent, Multiscale, and Multiplex Networks*, *Science* **328**, 876 (2010).
- [66] S. L. Braunstein, S. Ghosh, and S. Severini, *The Laplacian of a Graph as a Density Matrix: A Basic Combinatorial Approach to Separability of Mixed States*, *Ann. Comb.* **10**, 291 (2006).
- [67] H. Bamdad, F. Ashraf, and I. Gutman, *Lower Bounds for Estrada Index and Laplacian Estrada Index*, *Appl. Math. Lett.* **23**, 739 (2010).
- [68] I. Gutman, H. Deng, and S. Radenković, *Selected Topics on Applications of Graph Spectra* (Matematički Institut SANU, Belgrade, 2011), pp. 155–174.
- [69] E. Estrada and N. Hatano, *Statistical-Mechanical Approach to Subgraph Centrality in Complex Networks*, *Chem. Phys. Lett.* **439**, 247 (2007).
- [70] P. Erdős and A. Rényi, *On Random Graphs I*, *Publ. Math. Debrecen* **6**, 290 (1959).
- [71] J. Eisert, M. Cramer, and M. B. Plenio, *Colloquium: Area Laws for the Entanglement Entropy*, *Rev. Mod. Phys.* **82**, 277 (2010).
- [72] F. Franchini, A. R. Its, and V. E. Korepin, *Rényi Entropy of the XY Spin Chain*, *J. Phys. A* **41**, 025302 (2008).
- [73] J. Biamonte, V. Bergholm, and M. Lanzagorta, *Tensor Network Methods for Invariant Theory*, *J. Phys. A* **46**, 475301 (2013).
- [74] J. C. Baez, *Rényi Entropy and Free Energy*, arXiv: 1102.2098.
- [75] B. Schumacher and M. D. Westmoreland, *Relative Entropy in Quantum Information Theory*, *Contemp. Math.* **305**, 265 (2002).
- [76] A. Majtey, P. Lamberti, and D. Prato, *Jensen-Shannon Divergence as a Measure of Distinguishability between Mixed Quantum States*, *Phys. Rev. A* **72**, 052310 (2005).
- [77] J. Briët and P. Harremoës, *Properties of Classical and Quantum Jensen-Shannon Divergence*, *Phys. Rev. A* **79**, 052311 (2009).
- [78] P. Lamberti, A. Majtey, A. Borrás, M. Casas, and A. Plastino, *Metric Character of the Quantum Jensen-Shannon Divergence*, *Phys. Rev. A* **77**, 052311 (2008).
- [79] M. Gerlach, F. Font-Clos, and E. G. Altmann, *Similarity of Symbol Frequency Distributions with Heavy Tails*, *Phys. Rev. X* **6**, 021009 (2016).
- [80] D. Bures, *An Extension of Kakutani's Theorem on Infinite Product Measures to the Tensor Product of Semifinite  $w^*$ -Algebras*, *Trans. Am. Math. Soc.* **135**, 199 (1969).
- [81] C. Helstrom, *Minimum Mean-Squared Error of Estimates in Quantum Statistics*, *Phys. Lett. A* **25**, 101 (1967).
- [82] M. Hübner, *Explicit Computation of the Bures Distance for Density Matrices*, *Phys. Lett. A* **163**, 239 (1992).
- [83] H. Akaike, in *Proceedings of the Second International Symposium on Information Theory, Budapest, 1973*, edited by B. N. Petrov and F. Csaki, pp. 267–281.
- [84] M. A. Pitt, I. J. Myung, and S. Zhang, *Toward a Method of Selecting among Computational Models of Cognition*, *Psychol. Rev.* **109**, 472 (2002).
- [85] J. J. Rissanen, *Fisher Information and Stochastic Complexity*, *IEEE Trans. Inf. Theory* **42**, 40 (1996).
- [86] S.-i. Amari and H. Nagaoka, *Methods of Information Geometry* (American Mathematical Society, Providence, 2007), Vol. 191.
- [87] J. Rissanen, *Modeling by Shortest Data Description*, *Automatica* **14**, 465 (1978).



- [88] P. W. Holland, K. B. Laskey, and S. Leinhardt, *Stochastic Blockmodels: First Steps*, *Soc. Networks* **5**, 109 (1983).
- [89] M. De Domenico, S. Sasai, and A. Arenas, *Mapping Multiplex Hubs in Human Functional Brain Network*, *Front. Neurosci.* **10**, 326 (2016).
- [90] J. Iacovacci, Z. Wu, and G. Bianconi, *Mesoscopic Structures Reveal the Network between the Layers of Multiplex Data Sets*, *Phys. Rev. E* **92**, 042806 (2015).
- [91] N. Stanley, S. Shai, D. Taylor, and P. Mucha, *Clustering Network Layers with the Strata Multilayer Stochastic Block Model*, *IEEE Trans. Netw. Sci. Eng.* **3**, 95 (2016).
- [92] D. Taylor, S. Shai, N. Stanley, and P. J. Mucha, *Enhanced Detectability of Community Structure in Multilayer Networks through Layer Aggregation*, *Phys. Rev. Lett.* **116**, 228301 (2016).
- [93] P. Bientinesi, I. S. Dhillon, and R. A. Van De Geijn, *A Parallel Eigensolver for Dense Symmetric Matrices Based on Multiple Relatively Robust Representations*, *SIAM J. Sci. Comput.* **27**, 43 (2005).
- [94] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, *Asymptotic Analysis of the Stochastic Block Model for Modular Networks and Its Algorithmic Applications*, *Phys. Rev. E* **84**, 066106 (2011).
- [95] R. Guimerà and M. Sales-Pardo, *Missing and Spurious Interactions and the Reconstruction of Complex Networks*, *Proc. Natl. Acad. Sci.* **106**, 22073 (2009).
- [96] B. Karrer and M. E. Newman, *Stochastic Block Models and Community Structure in Networks*, *Phys. Rev. E* **83**, 016107 (2011).
- [97] M. E. J. Newman and A. Clauset, *Structure and Inference in Annotated Networks*, *Nat. Commun.* **7**, 11863 (2016).
- [98] T. P. Peixoto, *Parsimonious Module Inference in Large Networks*, *Phys. Rev. Lett.* **110**, 148701 (2013).
- [99] T. P. Peixoto, *Model Selection and Hypothesis Testing for Large-Scale Network Models with Overlapping Groups*, *Phys. Rev. X* **5**, 011033 (2015).
- [100] T. P. Peixoto and S. Bornholdt, *Evolution of Robust Network Topologies: Emergence of Central Backbones*, *Phys. Rev. Lett.* **109**, 118703 (2012).
- [101] L. Zdeborová and F. Krzakala, *Statistical Physics of Inference: Thresholds and Algorithms*, *Adv. Phys.* **65**, 453 (2016).
- [102] H. Jeffreys, *An Invariant Form for the Prior Probability in Estimation Problems*, *Proc. R. Soc. A* **186**, 453 (1946).
- [103] C. R. Rao, *Information and Accuracy Attainable in the Estimation of Statistical Parameters*, *Bull. Calcutta Math. Soc.* **37**, 81 (1945).
- [104] S.-I. Amari, *Information Geometry of the EM and em Algorithms for Neural Networks*, *Neural Netw.* **8**, 1379 (1995).
- [105] S.-I. Amari, *Information Geometry on Hierarchy of Probability Distributions*, *IEEE Trans. Inf. Theory* **47**, 1701 (2001).
- [106] S. L. Braunstein and C. M. Caves, *Statistical Distance and the Geometry of Quantum States*, *Phys. Rev. Lett.* **72**, 3439 (1994).
- [107] P. Hauke, M. Heyl, L. Tagliacozzo, and P. Zoller, *Measuring Multipartite Entanglement through Dynamic Susceptibilities*, *Nat. Phys.* **12**, 778 (2016).
- [108] M. M. Rams and B. Damski, *Quantum Fidelity in the Thermodynamic Limit*, *Phys. Rev. Lett.* **106**, 055701 (2011).
- [109] A. T. Rezakhani, D. F. Abasto, D. A. Lidar, and P. Zanardi, *Intrinsic Geometry of Quantum Adiabatic Evolution and Quantum Phase Transitions*, *Phys. Rev. A* **82**, 012321 (2010).
- [110] H. Strobel, W. Muessel, D. Linnemann, T. Zibold, D. B. Hume, L. Pezzè, A. Smerzi, and M. K. Oberthaler, *Fisher Information and Entanglement of Non-Gaussian Spin States*, *Science* **345**, 424 (2014).
- [111] P. Zanardi, P. Giorda, and M. Cozzini, *Information-Theoretic Differential Geometry of Quantum Phase Transitions*, *Phys. Rev. Lett.* **99**, 100603 (2007).
- [112] T. Manke, L. Demetrius, and M. Vingron, *Lethality and Entropy of Protein Interaction Networks*, *Genome Inform.* **16**, 159 (2005).
- [113] J. West, G. Bianconi, S. Severini, and A. E. Teschendorff, *Differential Network Entropy Reveals Cancer System Hallmarks*, *Sci. Rep.* **2**, 802 (2012).
- [114] G. Menichetti, G. Bianconi, G. Castellani, E. Giampieri, and D. Remondini, *Multiscale Characterization of Ageing and Cancer Progression by a Novel Network Entropy Measure*, *Mol. Biosyst.* **11**, 1824 (2015).
- [115] C. E. Granade, C. Ferrie, N. Wiebe, and D. G. Cory, *Robust Online Hamiltonian Learning*, *New J. Phys.* **14**, 103013 (2012).
- [116] N. Wiebe, C. Granade, C. Ferrie, and D. G. Cory, *Hamiltonian Learning and Certification Using Quantum Resources*, *Phys. Rev. Lett.* **112**, 190501 (2014).
- [117] A. Monras and F. Illuminati, *Information Geometry of Gaussian Channels*, *Phys. Rev. A* **81**, 062326 (2010).
- [118] L. Jing, J. Xiao-Xing, Z. Wei, and W. Xiao-Guang, *Quantum Fisher Information for Density Matrices with Arbitrary Ranks*, *Commun. Theor. Phys.* **61**, 115 (2014).