Generalized network structures: The configuration model and the canonical ensemble of simplicial complexes

Owen T. Courtney and Ginestra Bianconi

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

Simplicial complexes are generalized network structures able to encode interactions occurring between more than two nodes. Simplicial complexes describe a large variety of complex interacting systems ranging from brain networks to social and collaboration networks. Here we characterize the structure of simplicial complexes using their *generalized degrees* that capture fundamental properties of one, two, three, or more linked nodes. Moreover, we introduce the configuration model and the canonical ensemble of simplicial complexes, enforcing, respectively, the sequence of generalized degrees of the nodes and the sequence of the expected generalized degrees of the nodes. We evaluate the entropy of these ensembles, finding the asymptotic expression for the number of simplicial complexes in the configuration model. We provide the algorithms for the construction of simplicial complexes belonging to the configuration model and the canonical ensemble of simplicial complexes. We give an expression for the structural cutoff of simplicial complexes that for simplicial complexes of dimension $d = 1$ reduces to the structural cutoff of simple networks. Finally, we provide a numerical analysis of the natural correlations emerging in the configuration model of simplicial complexes without structural cutoff.

I. INTRODUCTION

Network theory has been successful over the past 15 years in characterizing social, technological, and biological networks. Nevertheless, the increasingly large data sets available in the field require the development of more sophisticated models of networks [1], such as multilayer networks [2,3] and generalized network structures [4,5]. In particular, a wide variety of networks, including brain networks [6,7], social and collaboration networks [8], immune networks [9], tagged social networks [4,5], and "folksonomies" [10,11], can be modeled by simplicial complexes [12–16]. Therefore, progress in understanding and modeling simplicial complexes has a variety of applications, ranging from brain research and data mining [6,17,18], to recommendation algorithms [19], characterization of dynamical processes [20], and inference of missing links [21].

Simplicial complexes are a generalization of networks constructed using not only nodes and links (that are, respectively, simplices of dimension zero and one) but also using triangles (simplices of dimension $d = 2$), tetrahedra (simplices of dimension $d = 3$), and higher-dimensional simplices. Using a theoretical physics terminology, simplicial complexes describe the *many-body* interactions between two or more nodes.

Simplicial complexes are emerging as a tool to describe complex networks with large clustering coefficient and abundant number of short loops that are not easily treatable by traditional statistical mechanics approaches. The presence of many short loops in real network datasets has often been recognized as a signature of a hidden geometry of networks [22,23]. Simplicial complexes are ideal mathematical objects for discretizing geometry as is demonstrated by their wide use in the context of quantum gravity [24–27] and therefore they can also open scenarios in uncovering the hidden geometry of complex networks.

Finally, simplicial complexes constitute the network-like structure that allows for the topological analysis of network datasets. The area of network topology is currently the subject of increasing interest, with recent investigations characterizing brain networks and network dynamics [6,7,17,18,20] providing results so far unobtainable through other network approaches.

For all these reasons it has become necessary to build null models for simplicial complexes using equilibrium and nonequilibrium approaches. Interestingly, extending our knowledge of static and growing network models [28–39] to simplicial complexes might reveal the role of the dimensionality of simplicial complexes in determining their structure.

Recently, a framework for nonequilibrium growing simplicial complexes has been formulated [12–15]. This framework is able to generate in one limit complex manifolds of dimension d and in another limit complex networks growing with preferential attachment. Interestingly, it has been observed that for dimension $d > 2$ growing manifolds are scale-free, because the increase of the dimensionality of simplicial complexes over $d = 2$ allows for the emergence of an efficient preferential attachment [14]. Interestingly, in this context it has also been shown that simplicial complexes growing by uniform attachment of simplices generate scale-free networks for $d \geqslant 2$ [15].

The formulation of equilibrium models of simplicial complexes is currently a hot topic in graph theory and pure mathematics [40–42]. Recently, exponential random simplicial complexes have also been attracting the attention of physicists and network scientists [16].

Here we develop an equilibrium statistical mechanics approach for simplicial complexes of dimension d. In particular we consider simplicial complexes formed exclusively by d-dimensional simplices. We characterize their structure with the generalized degree introduced in Refs. [13–15] and defined as the number of d-dimensional simplices incident to a given δ-dimensional face. Moreover, we treat in detail the configuration model and the canonical ensemble of simplicial complexes, respectively, with given generalized degree of the nodes and with expected generalized degrees of the nodes. The configuration model for simplicial complexes generalizes

the configuration model for simple networks [29–31] and the hypergraph model proposed in Ref. [4]. The canonical ensemble is instead to be related to exponential random simplicial complexes [16].

These ensembles can be treated using statistical mechanics arguments that are able to characterize their relation. Already in the context of simple networks one can distinguish between microcanonical and canonical conjugated network ensembles, which enforce, respectively, hard or soft constraints [29,32–35]. For example, the configuration model enforcing a given degree sequence and the exponential ensemble enforcing the expected degree sequence are, respectively, the micro-canonical and the canonical conjugated network ensembles. Similarly, here we show that the configuration model of sim-plicial complexes is the microcanonical ensemble conjugated to the canonical ensemble given by the exponential random simplicial complex. Interestingly, here we show that the two ensembles treated in this paper enforce an extensive number of constraints and, therefore, as already noted in the context of simple networks [29], they are not asymptotically equivalent.

The entropy of these ensembles, which has a number of ap-plications in network analysis and network inference [43,44], is here calculated analytically. From the entropy of the config-uration model of simplicial complexes the asymptotic combi-natorial formula for the number of simplicial complexes in the ensemble is derived. This formula generalizes the Canfield-Bender formula for the number of networks in the sparse configuration model [45]. When characterizing the properties of these ensembles, a special role is played by their structural cutoff that is the maximum generalized degree that guarantees the absence of correlations between the generalized degrees of the nodes in the simplicial complex. In any simplicial complex of dimension $d > 1$, the structural cutoff is larger than the structural cutoff of simple networks [46]. In the absence of the structural cutoff, simplicial complexes show relevant degree correlations analyzed here by numerical simulations. These results extend the known results observed in the canonical ensemble of simple networks [36].

The paper is structured as follows: in Sec. II we introduce simplicial complexes and the generalized degree of their nodes; in Sec. III we treat the canonical ensemble of simplicial complexes enforcing a given sequence of expected generalized degrees of the nodes; in Sec. IV we treat by statistical mechan-ics methods the configuration model of simplicial complexes with given sequence of generalized degrees of the nodes; in Sec. V we discuss the natural correlations observed in our numerical realizations of the the configuration model of sim-plicial complexes; finally, in Sec. VI we give the conclusions.

II. SIMPLICIAL COMPLEXES AND GENERALIZED DEGREES

A. Simplicial complexes of general dimension d

A d-dimensional simplex is formed by a set of $(d + 1)$ interacting nodes and includes all the subsets of $\delta + 1$ nodes (with $\delta < d$), which are called the δ-dimensional faces of the simplex.

A simplicial complex of dimension d is formed by simplices of dimension at most equal to d glued along their faces.

As mathematical objects simplicial complexes are distinct from hypergraphs [4,5], the difference being that simplicial complexes include all the subsets of a given simplex. Never-theless, in most of the interesting network science applications the terms simplicial complex and hypergraph might be used to indicate the same type of network data.

Here we consider d-dimensional simplicial complexes of N nodes formed exclusively by d-dimensional simplices. We indicate with $\mathcal{Q}_d(N)$ the set of all possible and distinct d-dimensional simplices in a d-dimensional simplicial complex of N nodes, while we indicate with $\mathcal{S}_{d,\delta}$ the set of all δ-dimensional simplices present in a given d-dimensional simplicial complex. The simplicial complexes that we consider in this paper are fully identified once the adjacency tensor **a** is fully specified. The adjacency tensor **a** has elements $a_\alpha = 0, 1$, indicating for each possible d-dimensional simplex $\alpha \in \mathcal{Q}_d(N)$ if the simplex is present ($a_\alpha = 1$) or absent ($a_\alpha = 0$) in the simplicial complex, i.e.,

$$a_\alpha = \begin{cases} 1 & \text{if } \alpha \in \mathcal{S}_{d,d} \\ 0 & \text{otherwise} \end{cases}.$$

The *generalized degrees* [14,15] are relevant structural prop-erties of simplicial complexes. The generalized degree $k_{d,\delta}(\alpha)$ of a δ-dimensional face (or δ-face) α of the d-dimensional simplicial complex quantifies the number of d-dimensional simplices incident to the δ-face α. The generalized degree $k_{d,\delta}(\alpha)$ can be defined in terms of the adjacency tensor **a** as

$$k_{d,\delta}(\alpha) = \sum_{\alpha' \in \mathcal{Q}_d(N)|\alpha' \supseteq \alpha} a_{\alpha'}. \tag{1}$$

The generalized degrees are not independent of each other. In fact, the generalized degree of a δ-face α is related to the generalized degree of the δ'-dimensional faces incident to it, with $\delta' > \delta$, by the simple combinatorial relation,

$$k_{d,\delta}(\alpha) = \frac{1}{\binom{d-\delta}{\delta'-\delta}} \sum_{\alpha \in \mathcal{S}_{d,\delta'}|\alpha' \supseteq \alpha} k_{d,\delta'}(\alpha'), \tag{2}$$

Moreover, since every d-dimensional simplex belongs to $\binom{d+1}{\delta+1}$ δ-dimensional faces, in a simplicial complex with M d-dimensional simplices we have

$$\sum_{\alpha \in \mathcal{S}_{d,\delta}} k_{d,\delta}(\alpha) = \binom{d+1}{\delta+1} M. \tag{3}$$

In this paper we focus specifically on the generalized degree of the nodes $r = 1, 2, \ldots N$ given by

$$k_{d,0}(r) = \sum_{\alpha' \in \mathcal{Q}_d(N)|\alpha' \supset r} a_{\alpha'}. \tag{4}$$

The generalized degree of the node indicates the number of d-dimensional simplices incident to each node r. Clearly, since the simplicial complexes under investigation are only formed by d-dimensional simplices, the generalized degree of the nodes satisfy

$$\sum_{r=1}^{N} k_{d,0}(r) = (d+1)M, \tag{5}$$

where M are the number of d-dimensional simplices in the simplicial complex. The generalized degree $k_{d,0}(r)$ of the nodes will play a crucial role in this paper because we will discuss the properties of the configuration model with given generalized degree sequence of the nodes. In the subsequent sections we will focus on the configuration model and the canonical ensemble for simplicial complexes enforcing, respectively, a given sequence of the generalized degree of the nodes and given sequence of the expected generalized degree of the nodes. We will always consider networks in which the number of simplices M is of the same order of magnitude of the number of nodes,

$$M \propto N, \tag{6}$$

which is the relevant regime for most of the applications to complex networks.

Before discussing the properties of these ensembles, in the following paragraphs we will characterize simplicial complexes of dimension $d = 1, 2$ using the generalized degrees of their δ-faces (for examples of simplicial complexes in dimension $d = 1, 2$; see Fig. 1). The extension to higher simplicial complexes is straightforward.

B. Case of a simplicial complex of dimension $d = 1$

Simplicial complexes of dimension $d = 1$ are formed exclusively by nodes and links (which are the 1-dimensional simplices). The adjacency tensor of the 1-dimensional simplicial complex is nothing else than the adjacency matrix $\{a_{rm}\}$, with elements a_{rm} indicating if the link (r, m) is present or not in the network. In this case the generalized degree $k_{1,0}(r)$ of the nodes, simply indicate the number of links incident to the node, i.e., its degree. In fact, we have

$$k_{1,0}(r) = \sum_{m=1}^{N} a_{rm}. \tag{7}$$

C. Case of a simplicial complex of dimension $d = 2$

Here we consider the case of a simplicial complex of dimension $d = 2$ characterizing interactions occurring between three nodes, i.e., a simplicial complex formed exclusively by triangles. We assume that the number of nodes in the simplicial complex is N. This simplicial complex is determined by the
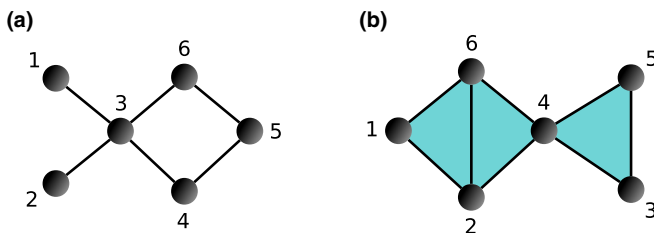

FIG. 1. Examples of simplicial complexes of dimension $d = 1$ (panel A) and $d = 2$ (panel B) are shown. Simplicial complexes of dimension $d = 1$ are simple networks. Simplicial complexes of dimension $d \geqslant 2$ characterize interactions occurring between more than two nodes (specifically interactions occurring between $d + 1$ nodes).

adjacency tensor $\{a_{rmn}\}$ of elements $a_{rmn} = 1$ if the nodes (r, m, n) are linked by a triangle, and $a_{rmn} = 0$ if the nodes (r, m, n) are not connected by a triangle. The generalized degree $k_{2,0}(r)$ of node r is given by

$$k_{2,0}(r) = \sum_{m<n} a_{rmn}, \tag{8}$$

while the generalized degree $k_{2,1}(r, m)$ of a link (r, m) is given by

$$k_{2,1}(r, m) = \sum_{n} a_{rmn}. \tag{9}$$

The generalized degree $k_{2,0}(r)$ of node r indicates the number of triangles incident to it, while the generalized degree $k_{2,1}(r, m)$ of the link (r, m) indicates the number of triangles incident to the link. The generalized degree of the nodes is related to the generalized degree of the links. In fact it is easy to see that

$$k_{2,0}(r) = \sum_{m<n} a_{rmn} = \frac{1}{2} \sum_{m,n} a_{rmn} = \frac{1}{2} \sum_{m} k_{2,1}(r, m). \tag{10}$$

Since each triangle is incident to three nodes, we have

$$\sum_{r=1}^{N} k_{d,0}(r) = 3M, \tag{11}$$

where M are the number of d-dimensional simplices in the simplicial complex.

III. CANONICAL ENSEMBLE OF SIMPLICIAL COMPLEXES

A. Canonical ensemble with given sequence of expected generalized degree of the nodes

In this section we discuss the canonical ensemble of simplicial complexes (also called the exponential random simplicial complex) with given sequence of expected generalized degree of the nodes. In this ensemble of simplicial complexes each simplicial complex G is assigned a probability $P(G)$. The entropy S of the ensemble evaluates the typical number of simplicial complexes belonging to the ensemble and is given by

$$S = -\sum_{G} P(G) \ln P(G), \tag{12}$$

where the sum is extended to all simplicial complexes G under consideration, or equivalently, over all adjacency tensors $\mathbf{a}$.

The canonical ensemble is the least-biased ensemble of simplicial complexes that satisfies the constraints,

$$\overline{k_r} = \overline{k_{d,0}(r)} = \sum_{G} P(G) \sum_{\alpha \in \mathcal{Q}_d(N)|r \subset \alpha} a_{\alpha}. \tag{13}$$

The canonical ensemble is the maximum entropy ensemble satisfying the constraints in Eq. (13). Therefore, in order to derive the probability $P(G)$ of a simplicial complex G in the

canonical ensemble, we maximize the functional $\mathcal{F}$ given by

$$\mathcal{F} = S + \sum_{r=1}^{N} \lambda_r \left[\overline{k_r} - \sum_G P(G) k_{d,0}(r) \right]$$
$$+ \mu \left[1 - \sum_G P(G) \right], \qquad (14)$$

where we have introduced, by means of the N Lagrangian multipliers λ_r, the constraints in Eq. (13), and by means of the Lagrangian multiplier μ the normalization constraint for the probability $P(G)$. Maximizing $\mathcal{F}$ with respect to $P(G)$, we obtain that the canonical ensemble of simplicial complexes enforcing a given sequence of expected generalized degrees of the nodes $\{\overline{k_r}\}$ has probability given by

$$P(G) = \frac{1}{Z} e^{-\sum_r \lambda_r k_{d,0}(r)}, \qquad (15)$$

where $k_{d,0}(r)$ is given by Eq. (4), and the normalization constant Z is given by

$$Z = \prod_{\alpha \in \mathcal{Q}_d(N)} [1 + e^{-\sum_{r \subset \alpha} \lambda_r}]. \qquad (16)$$

The Lagrangian multipliers $\{\lambda_r\}$ occurring in Eq. (15) are fixed by the Eqs. (13) and (15). Substituting the expression for $P(G)$ in Eq. (15) into Eq. (13), we get

$$\overline{k_r} = \sum_{\alpha \in \mathcal{Q}_d(N) | r \subset \alpha} \frac{e^{-\sum_{m \subset \alpha} \lambda_m}}{1 + e^{-\sum_{m \subset \alpha} \lambda_m}}. \qquad (17)$$

The probability p_α that each d-dimensional simplex $\alpha \in \mathcal{Q}_d(N)$ is in the simplicial complex is given by

$$p_\alpha = \sum_G P(G) a_\alpha = \frac{e^{-\sum_{r \subset \alpha} \lambda_r}}{1 + e^{-\sum_{r \subset \alpha} \lambda_r}}. \qquad (18)$$

Interestingly, for this ensemble, the probability $P(G)$ can be written as a product of the marginal probabilities for the individual d-dimensional simplices p_α, i.e.,

$$P(G) = \prod_{\alpha \in \mathcal{Q}_d(N)} \left[p_\alpha^{a_\alpha} (1 - p_\alpha)^{1 - a_\alpha} \right]. \qquad (19)$$

Consequently, the entropy S can be written as

$$S = - \sum_{\alpha \in \mathcal{Q}_d(N)} [p_\alpha \ln p_\alpha + (1 - p_\alpha) \ln(1 - p_\alpha)]. \qquad (20)$$

B. The canonical ensemble of simplicial complexes with structural cutoff

As long as the maximum generalized degree of the nodes is smaller than the structural cutoff, the probabilities p_α can be expressed as the normalized product of the generalized degrees of the nodes belonging to α. In fact, assuming $e^{-\lambda_r} \ll 1$, the probability p_α given by Eq. (18) can be approximated by

$$p_\alpha \simeq \prod_{r \subset \alpha, \alpha \in \mathcal{Q}_d(N)} e^{-\lambda_r}. \qquad (21)$$

Equation (17) can now be simplified and rearranged to give an explicit expression for $e^{-\lambda_r}$ in terms of k_r and the other Lagrangian multipliers:

$$e^{-\lambda_r} = \overline{k_r} \frac{d!}{(\sum_m e^{-\lambda_m})^d}. \qquad (22)$$

Note that in this last expression we made the following approximation

$$\sum_{m_1 < m_2 < \ldots < m_{d+1}} \prod_{j=1}^{d+1} e^{-\lambda_{m_j}} \simeq \frac{1}{d!} \left(\sum_m e^{-\lambda_m} \right)^d, \qquad (23)$$

valid in the limit in which the number of nodes N is large, i.e., $N \gg 1$ and $e^{-\lambda_r} \ll 1$. Summing over all the nodes of the simplicial complex, we get

$$\sum_r e^{-\lambda_r} = (\langle \overline{k} \rangle N d!)^{1/(d+1)}. \qquad (24)$$

Finally, combining Eqs. (22) and (24), we get

$$e^{-\lambda_r} = \overline{k_r} \left[\frac{d!}{(\langle \overline{k} \rangle N)^d} \right]^{1/(d+1)}. \qquad (25)$$

Using this result we get the simplified expression for the probability p_α of the d-dimensional simplex α, given by

$$p_\alpha = d! \frac{\prod_{r \subset \alpha} \overline{k_r}}{(\langle \overline{k} \rangle N)^d}, \qquad (26)$$

where $\alpha \in \mathcal{Q}_d(N)$. This expression is valid as long as $e^{-\lambda_r}$ [given by Eq. (25)] satisfies the hypothesis $e^{-\lambda_r} \ll 1$. This implies that the maximum generalized degree of the nodes $K_{\max}$ should be much smaller than the structural cutoff K_d for simplicial complexes; i.e.,

$$K_{\max} \ll K_d = \left[\frac{(\langle \overline{k} \rangle N)^d}{d!} \right]^{1/(d+1)}. \qquad (27)$$

Interestingly, the cutoff K_d for the present ensemble of simplicial complexes scales like $N^{d/(d+1)}$; i.e., it is increasing with an exponent that is larger for larger dimensions d.

This regime is the regime in which there are no correlations between the generalized degrees of the nodes. Moreover, in this regime only a few links can be incident to more than one d-dimensional simplex. In fact, given the expression for p_α provided by Eq. (26), it is possible to evaluate in this ensemble the expected generalized degree of the link $\overline{k_{d,2}(r,m)}$ for $d > 2$. This is given by

$$\overline{k_{d,1}(r,m)} = \sum_{\alpha | (r,m) \subset \alpha} p_\alpha = d \frac{\overline{k_r} \, \overline{k_m}}{\langle \overline{k} \rangle N}. \qquad (28)$$

Therefore, only the pairs of nodes (r,m) with generalized degree of the nodes $\overline{k_r}, \overline{k_m} \gg N^{1/2}$ and $\overline{k_r}, \overline{k_m} \ll N^{d/(d+1)}$ are likely to be incident to more than one d-dimensional simplex.

C. The canonical ensemble of simplicial complexes of dimension $d = 1$

For $d = 1$ our construction of the canonical ensemble of simplicial complexes for networks reduces to the canonical ensemble (exponential ensemble) of networks [28] with given expected degree sequence. The probability $P(G)$ of a given one-dimensional simplicial complex (i.e., network) specified

by the adjacency tensor $\{a_{rm}\}$ is given by Eq. (15), given in this case by

$$P(G) = \frac{1}{Z} e^{-\sum_r \lambda_r k_{1,0}(r)}, \tag{29}$$

where $k_{1,0}(r)$ is given by Eq. (7) and the normalization constant Z is given by

$$Z = \prod_{r<m} (1 + e^{-\lambda_r - \lambda_m}). \tag{30}$$

The Lagrangian multipliers λ_r are fixed by the condition

$$\overline{k_r} = \overline{k_{1,0}(r)} = \sum_m p_{rm}, \tag{31}$$

with p_{rm} indicating the probability that the link between the nodes r,m is present in the network. The probability p_{rmn} are given by

$$p_{rm} = \frac{e^{-(\lambda_r + \lambda_m)}}{1 + e^{-(\lambda_r + \lambda_m)}}. \tag{32}$$

The probability $P(G)$ of a simplicial complex G in this canonical ensemble can be expressed as a product of the marginal probabilities for the individual links:

$$P(G) = \prod_{r<m} \left[p_{rm}^{a_{rm}} (1 - p_{rm})^{1-a_{rm}} \right]. \tag{33}$$

Therefore, the entropy S of the ensemble is given by

$$S = -\sum_{r<m} [p_{rm} \ln p_{rm} + (1 - p_{rm}) \ln(1 - p_{rm})]. \tag{34}$$

Finally, in the presence of the structural cutoff on the generalized degree of the nodes, i.e., if the maximal generalized degree of the nodes $K_{\max}$ satisfies

$$K_{\max} \ll K_1 = (\langle \overline{k} \rangle N)^{1/2}, \tag{35}$$

the probabilities p_{rm} take a simple factorized expression given by

$$p_{rm} = \frac{\overline{k_r}\, \overline{k_m}}{\langle \overline{k} \rangle N}. \tag{36}$$

We note here that the structural cutoff of simplicial complexes of dimension $d = 1$ given by Eq. (35) reduces to the structural cutoff of simple networks [46] as expected.

D. The canonical ensemble of simplicial complexes of dimension $d = 2$

In this subsection we summarize the results for the case of a canonical ensemble of two-dimensional simplicial complexes where we constrain the expected generalized degree of the nodes to be $k_{2,0}(r)$. The probability $P(G)$ of a given simplicial complex specified by the adjacency tensor $\{a_{rmn}\}$, which is given by Eq. (15) that reads for this case

$$P(G) = \frac{1}{Z} e^{-\sum_r \lambda_r k_{2,0}(r)}, \tag{37}$$

where $k_{2,0}(r)$ is given by Eq. (8) and the normalization constant Z is given by

$$Z = \prod_{r<m<n} (1 + e^{-\lambda_r - \lambda_m - \lambda_n}). \tag{38}$$

The Lagrangian multipliers λ_r are fixed by the condition

$$\overline{k_r} = \overline{k_{2,0}(r)} = \sum_{m<n} p_{rmn}, \tag{39}$$

with p_{rmn} indicating the probability that the triangle between the nodes r,m,n is present in the simplicial complex, which is given by

$$p_{rmn} = \frac{e^{-(\lambda_r + \lambda_m + \lambda_n)}}{1 + e^{-(\lambda_r + \lambda_m + \lambda_n)}}. \tag{40}$$

The probability $P(G)$ of a simplicial complex G in this canonical ensemble can be expressed as a product of the marginal probabilities for the individual triangles:

$$P(G) = \prod_{r<m<n} \left[p_{rmn}^{a_{rmn}} (1 - p_{rmn})^{1-a_{rmn}} \right]. \tag{41}$$

Therefore, the entropy S of the ensemble is given by

$$S = -\sum_{r<m<n} [p_{rmn} \ln p_{rmn} + (1 - p_{rmn}) \ln(1 - p_{rmn})]. \tag{42}$$

Finally, in the presence of the structural cutoff on the generalized degree of the nodes, i.e., if the maximal generalized degree of the nodes $K_{\max}$ satisfies

$$K_{\max} \ll K_2 = \left(\frac{\langle \overline{k} \rangle N}{\sqrt{2}} \right)^{2/3}, \tag{43}$$

then the probabilities p_{rmn} take a simple factorized expression given by

$$p_{rmn} = 2 \frac{\overline{k_r}\, \overline{k_m}\, \overline{k_n}}{(\langle \overline{k} \rangle N)^2}. \tag{44}$$

Here the structural cutoff K_2 scales like $N^{2/3}$. It is, therefore, much larger than the structural cutoff for simple networks.

We note that this model is to be related with the model of tagged social networks represented by hypergraphs presented in Refs. [4,5]. Nevertheless, it differs with respect to the cited work because in the present work the three nodes linked in a given two-dimensional simplex represent the same type of nodes. This difference is responsible for the factor two present in the righthand side of Eq. (44).

E. Generation of simplicial complexes by the canonical ensemble

For generating the canonical ensemble of d-dimensional simplicial complexes with expected generalized degree sequence of the nodes $\{\overline{k_r}\}$ with $r = 1, 2, \ldots N$, we propose the following algorithm:

(a) Calculate the probabilities p_α of any d-dimensional simplex $\alpha \in \mathcal{Q}_d(N)$ given by Eq. (18) in the absence of the structural cutoff K_d or by Eq. (26) in presence of the structural cutoff K_d.

(b) Draw every possible d-dimensional simplex $\alpha \in \mathcal{Q}_d(N)$ with probability p_α.

IV. THE CONFIGURATION MODEL
OF SIMPLICIAL COMPLEXES

A. The configuration model of simplicial complexes
with given generalized degree of the nodes

The configuration model of simplicial complexes with given sequence of the generalized degrees of the nodes $\{k_r\}$ is the ensemble in which we assign the same probability to each simplicial complex with the generalized degrees of the nodes satisfying $k_{d,0}(r) = k_r$ for every node r. The construction of simplicial complexes is allowed only if the generalized degree sequence of the nodes is graphical, i.e., if at least one simplicial complex can be constructed with it. For simple networks, i.e., for simplicial complexes of dimension $d = 1$, the conditions that a degree sequence must satisfy in order to be graphical have been fully identified [30,47]. For simplicial complexes we know that at least the generalized degree of the nodes must satisfy Eq. (5). In practice, it will often be useful to start from sequences of generalized degree of the nodes occurring in real data sets, which are by definition graphical. This will be recommended in order to construct a randomized simplicial complex that will provide a null model to the real dataset.

The configuration model enforcing a given graphical sequence of the generalized degrees of the nodes assigns to each d-dimensional simplicial complex G formed exclusively by d-dimensional simplexes the probability

$$P(G) = \frac{1}{\mathcal{N}} \prod_{r=1,N} \delta(k_r, k_{d,0}(r)). \tag{45}$$

Here $\mathcal{N}$ is the number of simplicial complexes with the given graphical sequence of generalized degree of the nodes $\{k_r\}$ given by

$$\mathcal{N} = \sum_G \prod_{r=1,N} \delta(k_r, k_{d,0}(r)). \tag{46}$$

In Fig. 2 we show how from a given graphical sequence of generalized degree of the nodes it is possible in general to construct different simplicial complexes.

B. Generation of the simplicial complexes
by the configuration model

In this subsection we generalize the algorithm for the configuration model of networks with given degree sequence to the configuration model of d-dimensional simplicial complexes with given sequence $\{k_r\}_{r \leqslant N}$ of the generalized degrees of the nodes. For describing this algorithm, we will use a set of M auxiliary *factor nodes* $\mu = 1, 2, \ldots, M$, with M satisfying

$$\sum_{r=1}^{N} k_r = (d+1)M. \tag{47}$$

The algorithm is described in Fig. 3 in the case $d = 2$ and proceeds as follows:

(i) Initially, k_r stubs are placed on each node $r = 1, 2, \ldots, N$. Additionally, $d+1$ stubs are placed on each auxiliary factor node $\mu = 1, 2, \ldots M$. Initially each stub is unmatched.

(ii) A set of $d+1$ unmatched random stubs of the nodes is chosen with uniform probability. Without losing
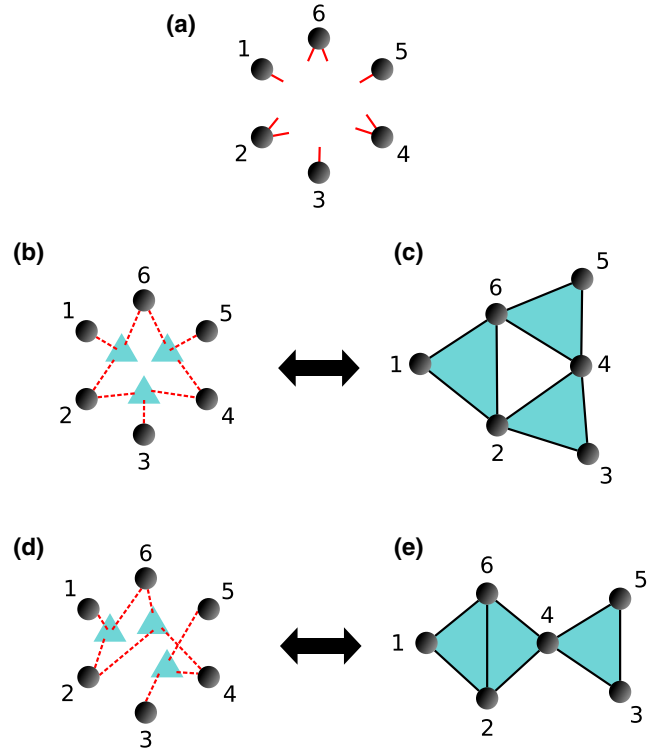

FIG. 2. The figure shows the construction of two different $d = 2$ dimensional simplicial complexes belonging to the same configuration model of simplicial complexes. In panel (a) the $N = 6$ nodes are shown together with stubs, indicating their generalized degree. In panel (b) triples of stubs are matched together to form two-dimensional simplices. In panel (c) the corresponding simplicial complex is visualized. In panels (d) and (e) a different matching of the stubs is shown together with its corresponding simplicial complex. As is evident from the figure, a given generalized degree sequence of the nodes can give rise to different simplicial complexes. The logarithm of the total number $\mathcal{N}$ of simplicial complexes that can be constructed from a given generalized degree sequence of the nodes is the Gibbs entropy Σ of the configuration model.

generality we assume that the stubs belong to the set of nodes $(r_1, r_2, \ldots, r_{d+1})$.

(iii) If the nodes $(r_1, r_2, \ldots, r_{d+1})$ are all distinct and no factor node μ is already matched with the set of nodes $(r_1, r_2, \ldots, r_{d+1})$, we match the $d+1$ stubs of an unmatched random factor node to the nodes $(r_1, r_2, \ldots, r_{d+1})$. Otherwise, we start again from step (i).

(iv) If all the stubs are matched we construct the simplicial complex by placing a simplex between the nodes connected to each auxiliary factor node.

In Fig. 3 we show an example of the possible matching of the stubs of nodes and factor nodes and the consequent construction of the simplicial complex.

Step (iii) rejects moves that are forbidden. These moves are described in Fig. 4. This rejection procedure guarantees that there are no spurious correlations in the structure of the simplicial complex, but for broad distribution of the generalized degrees of the nodes it might significantly slow down the algorithm. In the context of the configuration model, more sophisticated algorithms have been proposed in
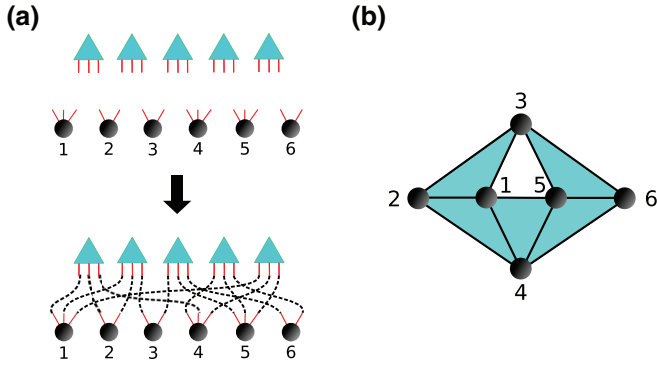
FIG. 3. A scheme representing the algorithm for the construction of the configuration model is shown for the case $d = 2$. Panel A represents steps (i), (ii), and (iii). To each node r with $r = 1, 2 \ldots, N = 6$ we assign k_r stubs. The nodes are represented with black circles. A set of M auxiliary factor nodes (cyan triangles) is considered. Each factor node has $d + 1$ stubs. Subsequently, an allowed matching of the stubs is found. Panel B shows how from the matching of the stubs we can construct a simplicial complex by adding a simplex between all of the nodes connected to a common factor node in panel A.

Refs. [30,31] and we believe that along these lines it could also be possible to optimize the code for the case of simplicial complexes in the future.

Here, when numerically implementing the algorithm (see Supplemental Material [48] for the codes generating random simplicial complexes in $d = 1, d = 2, d = 3$), we have chosen to allow a rejection of a small number n_F of forbidden moves. Therefore, we have modified the above algorithm by substituting step (iii) with:

(iii a) If the nodes $(r_1, r_2, \ldots, r_{d+1})$ are all distinct and no factor node μ is already matched with the set of nodes $(r_1, r_2, \ldots, r_{d+1})$, we match the $d + 1$ stubs of an unmatched random factor node to the nodes $(r_1, r_2, \ldots, r_{d+1})$.
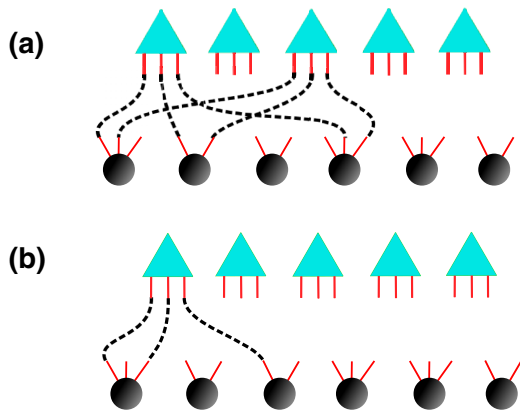

FIG. 4. Two examples of forbidden moves are shown. In panel A the same set of nodes $(r_1, r_2, \ldots, r_{d+1})$ is selected more than once to form a simplex. In panel B the set of nodes $(r_1, r_2, \ldots, r_{d+1})$ selected to form a simplex is not formed by $d + 1$ distinct nodes. Here the forbidden moves are shown for the configuration model of simplicial complexes of dimension $d = 2$.

(iii b) If the nodes $(r_1, r_2, \ldots, r_{d+1})$ are not all distinct or a factor node μ is already matched with the set of nodes $(r_1, r_2, \ldots, r_{d+1})$, we update a variable n_x that counts how many similar events have occurred so far. If $n_x \leqslant n_F$, we do not accept the move and we go back to step (ii); if $n_x > n_F$, we go back to step (i).

This algorithm reduces to the one described before when $n_F = 1$, and when $n_F \ll N$ it speed up significantly the code, without altering significantly the properties of the simplicial complexes.

C. Relation with bipartite network models

Bipartite networks are formed by a set of nodes $r = 1, 2, \ldots, N$ and a set of factor nodes (groups) $\mu = 1, 2, \ldots, P$ where links only join a node with a factor node (a group). A given bipartite network has adjacency matrix **A** with elements $A_{r,\mu} = 1$ if the node r belongs to group μ, and $A_{r,\mu} = 0$ otherwise. The bipartite network might, for example, describe a network formed by scientists (the nodes) and by scientific papers (the groups) where each paper is connected to the set of its authors. Similar models have been proposed for social networks [8] and for immune networks [9]. Each group μ of a bipartite network can be related to a simplicial complex constructed by joining all the nodes connected to a common factor node (group). In particular, if all the factor nodes have the degree equal to d, all these simplicial complexes are d-dimensional. Therefore, it is important to discuss here the relation between the configuration model for simplicial complexes and the ensemble of bipartite networks in which we fix the degree sequence of the nodes and of the factors nodes to be

$$\sum_\mu A_{r\mu} = k_r,$$

$$\sum_{r=1}^N A_{r\mu} = d + 1. \tag{48}$$

The differences between the configuration model for bipartite networks with constraints given by Eq. (48) and the configuration model for simplicial complexes are:

(1) In the bipartite network it is possible to observe more than one factor node connecting the same set of nodes.

(2) In the bipartite network the factor nodes are labeled.

An example illustrating these differences is that of a bipartite network between authors and papers coauthored by three authors and the corresponding simplicial complex describing the collaboration network between the authors. The difference between these two data sets is that bipartite networks distinguish between situations where three authors write only one or several papers together, and they also distinguish between papers with the same three authors (i.e., the papers are labeled). In contrast, simplicial complexes indicate only whether a given set of three authors have coauthored at least one paper together, independently on the paper title and content.

D. Canonical ensemble conjugated to the configuration model of simplicial complexes

The configuration model for simplicial complexes enforcing the generalized degree sequence of the nodes $\{k_r\}$ and the canonical ensemble of simplicial complexes enforcing the expected generalized degree of the nodes $\{\overline{k}_r\}$ are *conjugated network ensembles* when $\overline{k}_r = k_r$ for every node r. The configuration model can also be called the microcanonical ensemble conjugated to the canonical ensemble of simplicial complexes. This terminology is borrowed from statistical mechanics that treats ensembles of dynamical systems with given energy (microcanonical ensemble) or with given expected (average) energy (canonical ensemble). In statistical mechanics these two ensembles are thermodynamically equivalent; i.e., their statistical properties are the same when one considers systems formed by large number of particles as, for example, a gas of molecules. For network ensembles, the most fundamental example of conjugated microcanonical and canonical ensembles are the Erdös and Renyi random graphs in which we fix the total number of links (microcanonical ensemble) and the random graph in which we fix the average number of links (canonical ensemble). These network ensembles are equivalent in the thermodynamic limit like the microcanonical and the canonical ensemble of a gas of molecules. Nevertheless, in network theory it is often the case that we are interested in characterizing network ensembles with an extensive number of constraints like the ones in which we fix the degree sequence (microcanonical ensemble) or the expected degree sequence (canonical ensemble). In these cases we no longer observe the equivalence of the two conjugated ensembles [29,35]. In the following section we will provide evidence that the configuration model of simplicial complexes and its conjugated canonical ensemble of simplicial complexes are not asymptotically equivalent.

E. The entropy of the configuration model of simplicial complexes

The entropy evaluates the logarithm of the (typical) number of simplicial complexes in the ensemble. This quantity characterizes the complexity of the constraints or, in other words, *how complex are the simplicial complexes in the ensemble* [32]. In fact, an ensemble constructed from very stringent, complex constraints will give rise to few network realizations. Therefore, the entropy of network ensembles can be used in a variety of inference problems [43,44]. Moreover, the entropy of conjugated ensembles can indicate whether the two ensembles are asymptotically equivalent. In fact, if the entropy of two conjugated ensembles is not the same in the large network limit, the two ensembles are not asymptotically equivalent. In this paragraph we provide a summary of the results obtained for the entropy of the configuration model of simplicial complexes. The details of the derivations will be given in the Appendices.

The entropy Σ of the microcanonical ensemble defined by the configuration model of simplicial complexes with given sequence $\{k_r\}$ of generalized degree of the nodes is defined as the logarithm of the number of simplicial complexes belonging

to the ensemble, i.e.,

$$\Sigma = \ln \mathcal{N} = \ln \left[\sum_G \prod_r \delta(k_r, k_{d,0}(r)) \right]. \quad (49)$$

In fact, it can be easily shown that

$$\Sigma = \ln \mathcal{N} = -\sum_G P(G) \ln P(G), \quad (50)$$

where $P(G)$ is given by Eq. (45). To distinguish the entropy Σ from the entropy S of the canonical ensemble defined in Eq. (12), we call Σ the Gibbs entropy and S the Shannon entropy. When considering the entropies Σ and S of conjugated microcanonical (configuration model) and canonical ensemble, with $\overline{k}_r = k_r$, $\forall r$ we obtain (see Appendix A for details)

$$\Sigma = S - \Omega, \quad (51)$$

where Ω is the entropy of large deviation, which is the logarithm of the probability that in the canonical network model with expected generalized degree sequence $\{\overline{k}_r\}$ (with $\overline{k}_r = k_r$), the generalized degrees of the nodes take exactly the values $k_{d,0}(r) = k_r$. Therefore, Ω can be expressed as

$$\Omega = -\ln \left[\sum_G P(G) \prod_r \delta(k_r, k_{d,0}(r)) \right], \quad (52)$$

where

$$P(G) = \prod_{\alpha \in \mathcal{Q}_d(N)} \left[p_\alpha^{a_\alpha} (1 - p_\alpha)^{1-a_\alpha} \right], \quad (53)$$

with

$$p_\alpha = \frac{e^{-\sum_{r \subset \alpha} \lambda_r}}{1 + e^{-\sum_{r \subset \alpha} \lambda_r}}, \quad (54)$$

and

$$k_r = \sum_{\alpha | r \subset \alpha} p_\alpha. \quad (55)$$

Since Ω is nonnegative, Eq. (51) shows that the Gibbs entropy Σ is less than or equal to the Shannon entropy S, and when Ω is not negligible, the two entropies are not the same, indicating a nonequivalence of the microcanonical (configuration model) and the canonical ensemble for simplicial complexes.

For simplicial complexes with the structural cutoff, Ω given by Eq. (52) can be calculated using the saddle point approximation (see Appendix B for details), obtaining

$$\Omega = -\sum_{r=1}^N \ln[\pi_{k_r}(k_r)], \quad (56)$$

where $\pi_{k_r}(k_r)$ is the Poisson distribution with average k_r evaluated at k_r; i.e.,

$$\pi_{k_r}(k_r) = \frac{1}{k_r!} k_r^{k_r} e^{-k_r}. \quad (57)$$

This expression is easily interpreted. In fact, in the canonical ensemble the generalized degree of each node follows a Poisson distribution with average $\overline{k}_r = k_r$. Therefore, the probability that each of these generalized degrees takes exactly the value $k_{d,0}(r) = k_r$ is given by $\pi_{k_r}(k_r)$. We note here that

Ω given by Eq. (56) is an extensive quantity and therefore the Gibbs entropy Σ is significantly different from the Shannon entropy S of the canonical ensemble, implying that the two conjugated ensembles are not asymptotically equivalent.

F. The asymptotic formula for the number of simplicial complexes in the configuration model with structural cutoff

The Gibbs entropy Σ can be evaluated using Eq. (51) together with Eq. (56) and the expression given by Eq. (20) for the Shannon entropy, as long as we are in the presence of the structural cutoff K_d defined in Eq. (27), getting

$$\Sigma = -\sum_{\alpha \in \mathcal{Q}_d(N)} [p_\alpha \ln p_\alpha + (1 - p_\alpha) \ln(1 - p_\alpha)]$$
$$+ \sum_{r=1}^{N} \ln \frac{k_r^{k_r} e^{-k_r}}{k_r!}, \tag{58}$$

where in presence of the structural cutoff the probabilities p_α are given by Eq. (26) with $\overline{k_r} = k_r$ for every node r. Substituting the expression for p_α into Eq. (58) we get the asymptotic expression for the logarithm of the number of simplicial complexes $\mathcal{N}$ in the configuration model,

$$\Sigma = \ln \mathcal{N}$$
$$= \frac{d}{d+1} \ln(\langle k \rangle N)! - \sum_{r=1}^{N} \ln k_r! - \frac{\langle k \rangle N}{d+1} \ln d!$$
$$- \frac{d!}{2(d+1)(\langle k \rangle N)^{d-1}} \left(\frac{\langle k^2 \rangle}{\langle k \rangle} \right)^{d+1}. \tag{59}$$

Therefore, the asymptotic expression for the number $\mathcal{N}$ of simplicial complexes in the configuration model is given by

$$\mathcal{N} = \frac{[(\langle k \rangle N)!]^{d/(d+1)}}{\prod_{r=1}^{N} k_r!} \frac{1}{(d!)^{\langle k \rangle N/(d+1)}}$$
$$\times \exp \left[-\frac{d!}{2(d+1)(\langle k \rangle N)^{d-1}} \left(\frac{\langle k^2 \rangle}{\langle k \rangle} \right)^{d+1} + \mathcal{O}(\ln N) \right]. \tag{60}$$

This expression is the generalization of the Canfield-Bender formula [45] for the ensemble of networks with given degree sequence. In fact, for $d = 1$ it is reduced to the Canfield-Bender formula. Interestingly, we observe that the asymptotic number $\mathcal{N}$ of simplicial complexes in the configuration model depends on the distribution of the generalized degrees of the nodes and that this dependency remains important even for generalized degree sequences with the same average $\langle k \rangle$. This shows that the complexity of the ensemble depends strongly on the statistical characteristics of the generalized degree sequence. As observed in Ref. [32] in the context of simple networks, it can also be shown for simplicial complexes of dimension $d > 1$ that scale-free distributions of generalized degrees with the same average $\langle k \rangle$ but with decreasing power-law exponent γ correspond to more complex ensembles of simplicial complexes. In fact, they are characterized by a smaller entropy Σ and a smaller asymptotic number $\mathcal{N}$ of simplicial complexes.

G. Combinatorial arguments for Eq. (60)

The asymptotic combinatorial expression [Eq. (60) can be explained using combinatorial arguments, similar to the ones used to explain the Canfield-Bender formula in Ref. [33]. In fact, the factor

$$\frac{[(\langle k \rangle N)!]^{d/(d+1)}}{\prod_{r=1}^{N} k_r!} \frac{1}{(d!)^{\langle k \rangle N/(d+1)}} \tag{61}$$

counts all the possible combinations of the stubs of the nodes in groups of $d + 1$ stubs when we disregard forbidden moves. In other words, Eq. (61) counts all the possible matchings between the stubs of the nodes and the stubs of the factor nodes obtained by following the algorithm described in Sec. IV B, considering the fact that the factor nodes are not labeled and neglecting the occurrence of forbidden moves. In fact, if we want to construct a simplicial complex with a given sequence of the generalized degree of the nodes $\{k_r\}$, the first step is to take a stub of a node and match it with a stub of an unmatched factor node. Since the factor nodes are not labeled every unmatched factor node is equivalent, and therefore there is a unique way to match a given stub of the node with an unmatched unlabeled factor node. Subsequently, we proceed with matching the remaining d stubs of this newly matched factor node. In order to do this, we choose an unordered set of d of the remaining stubs of the nodes. We have $[\langle k \rangle N - 1][\langle k \rangle N - 2] \ldots [\langle k \rangle N - (d + 1) - 1]/d!$ ways to perform this step if we neglect forbidden moves. Once these stubs have been chosen, all the stubs of the first factor node are now matched to $d + 1$ stubs of the nodes. The remaining factor nodes are all unmatched. The next step is to take one of the remaining $\langle k \rangle N - (d + 1)$ stubs of the nodes and to match it to an arbitrary unmatched factor node. Since the unmatched factor nodes are all equivalent,
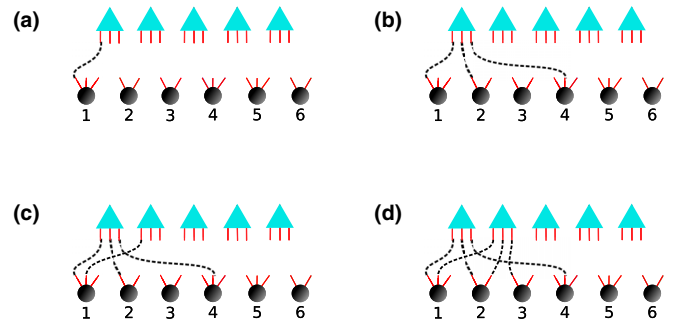

FIG. 5. The subsequent matching of the node stubs with the stubs of the factor nodes is shown here in order to justify Eq. (61) evaluating the asymptotic number of matchings in the absence of forbidden moves. First a random stub of a node is matched with a stub of an arbitrary unmatched factor node (panel A). Subsequently, all the remaining d stubs of the factor node are matched with stubs of the nodes (panel B). At this point the unmatched factor nodes are reduced by one. An unmatched node-stub is matched to a stub of an arbitrary unmatched factor node (panel C). Subsequently, all the remaining stubs of the second factor node are matched with stubs of the nodes (panel D). This procedure continues in the absence of forbidden moves, until all of the stubs of the nodes are matched with all the stubs of the factor nodes. By calculating the probability of these moves, as describe in Sec. IV G we can derive Eq. (61).
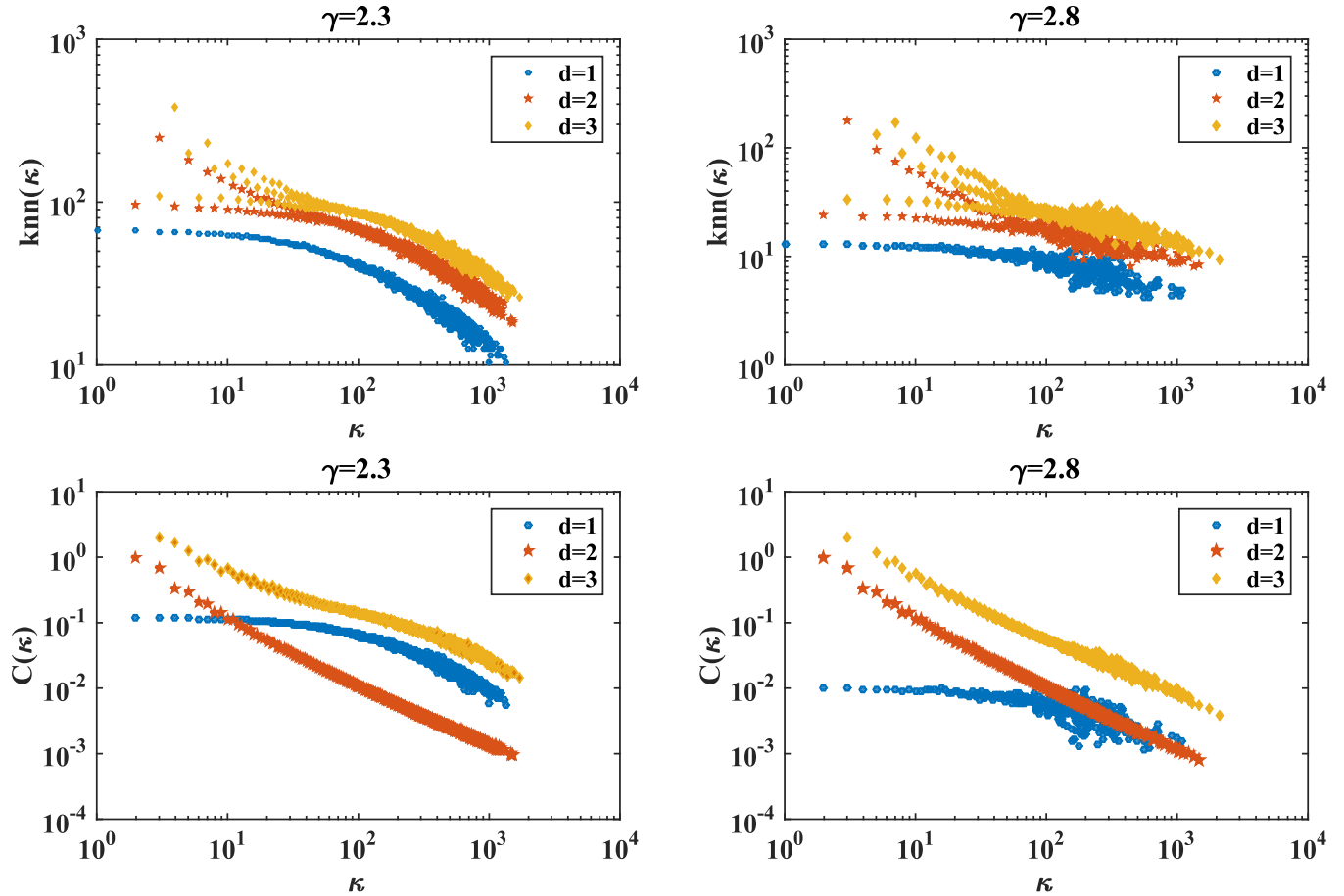
FIG. 6. The average degree $knn(\kappa)$ of the neighbors of the nodes of degree κ and the average clustering coefficient $C(\kappa)$ of the nodes of degree κ, for simplicial complexes of dimension $d = 1,2,3$ constructed according to the configuration model with distribution of the generalized degrees of the nodes $P_{d,0}(k)$ given by Eq. (65) and $\gamma = 2.3,2.8$. The simplicial complexes have $N = 10^4$ nodes, and $n_F = 70$. The data are averaged over 100 realizations.

there is a unique way to do this. Subsequently, we proceed as we have done previously and we find d remaining stubs of the nodes to be matched to the second factor node. There are $\langle k \rangle N - (d + 2)$ remaining stubs to choose from, so there are $[\langle k \rangle N - (d + 2)][\langle k \rangle N - (d + 3)] \ldots [\langle k \rangle N - 2(d + 1) - 1]/d!$ ways of selecting this unordered set of d stubs. For a visual representation of these steps, see Fig. 5. It is easy to see in this way that we can proceed by matching all the stubs of the nodes with the indistinguishable factor nodes in

$$\frac{(\langle k \rangle N - 1)!}{\prod_{s=1}^{\frac{\langle k \rangle N}{d+1}-1}[\langle k \rangle N - s(d+1)])} \frac{1}{(d!)^{\langle k \rangle N/(d+1)}} \qquad (62)$$

ways since $\langle k \rangle N/(d + 1) = M$ counts the number of factor nodes. As long as d is finite and $N \gg 1$, we can use the following approximation:

$$\frac{(\langle k \rangle N - 1)!}{\prod_{s=1}^{\frac{\langle k \rangle N}{d+1}-1}[\langle k \rangle N - s(d+1)])} \simeq [(\langle k \rangle N)!]^{\frac{d}{d+1}}, \qquad (63)$$

getting the asymptotic approximation for Eq. (62) given by

$$\frac{[(\langle k \rangle N)!]^{d/(d+1)}}{(d!)^{\langle k \rangle N/(d+1)}}. \qquad (64)$$

In order to find the number of distinct matchings given by Eq. (61) we need to observe that all the permutations of the stubs of each single node give equivalent matchings. We need, therefore, to divide the expression found in Eq. (64) by $\prod_{r=1,N} k_r!$ getting Eq. (61). Finally, the exponential term in Eq. (60) needs to be interpreted as the term that corrects for the forbidden matchings.

V. NATURAL CORRELATIONS OF THE CONFIGURATION MODEL OF SIMPLICIAL COMPLEXES

The configuration model of simplicial complexes without structural cutoff develops significant degree correlations. In order to characterize the degree correlations present in the simplicial complexes of different dimension d, we have considered the simplicial complexes constructed by the configuration model with scale-free distribution $P_{d,0}(k)$ of the generalized degree of the nodes $k_{d,0} = k$. The distribution $P_{d,0}(k)$ of the generalized degree of the nodes is given by

$$P_{d,0}(k) = Ck^{-\gamma}, \qquad (65)$$

with minimal generalized degree $m = 1$. The generated simplicial complexes can be analyzed by means of the

well established tools used in network theory. In particular, a network structure can be extracted from the simplicial complexes by assuming that two nodes are linked if they belong at least to a common simplex. We will call the adjacency matrix of this network $\hat{\mathbf{a}}$ and the degree of the generic node r in this network κ_r. The correlations existing in this network can be characterized by means of the average degree $knn(\kappa)$ of the neighbors of the nodes of degree κ, and the average clustering coefficient $C(\kappa)$ of the nodes of degree κ. These functions are plotted in Fig. 6 for simplicial complexes with generalized degree distribution $P_{d,0}(k)$ given by Eq. (65) and $\gamma = 2.3, 2.8$. These results show that natural degree correlations occur in these models. The average clustering coefficient $C(\kappa)$ increases with the increasing dimensionality d of the simplicial complex, and the shape of the function $C(\kappa)$ also is strongly dependent on the dimensionality d. On the contrary, $knn(\kappa)$ does not appear to change so dramatically with the dimensionality d of the simplicial complex.

VI. CONCLUSIONS

Simplicial complexes of dimension $d > 1$ encode information about interactions occurring between more than two nodes while simplicial complexes of dimension $d = 1$ are simple networks describing only pairwise interactions. As such, simplicial complexes are a generalization of network structures that can be extremely useful for analyzing a large variety of complex interacting systems ranging from brain networks to social networks. As approaches to data analysis of networked systems require the characterization of complex datasets in terms of simplicial complexes, building null models for these structures is increasingly important for the advance of the field. Here we have characterized the structural properties of simplicial complexes using the generalized degrees, which capture fundamental properties of their δ faces. We have fully investigated the configuration model for simplicial complexes with statistical mechanics techniques relating its properties with the ones of the conjugated canonical ensemble of simplicial complexes (also called the exponential random simplicial complex). The entropy of these ensembles is derived here with analytical techniques, opening the possibility to use this quantity as an information theory measure for inference problems on simplicial complexes. Additionally, we have found an expression for the structural cutoff of simplicial complexes that generalizes the structural cutoff of the configuration model of simple networks. Finally, we have provided algorithms for generating simplicial complexes belonging to the configuration model and the canonical ensembles studied in this paper, and we have numerically investigated the natural correlations emerging in these models.

In conclusion, we believe that this paper provides a full account of two of the most fundamental equilibrium models of simplicial complexes which can be used as null models for investigating the structure of simplicial complexes, or for studying dynamical processes. We believe that these models constitute only the first step in modeling simplicial complexes with equilibrium statistical mechanics tools and that our work will open perspectives for investigating a new generation of equilibrium models for simplicial complexes.

APPENDIX A: DERIVATION OF EQ. (51) RELATING THE ENTROPY Σ CONFIGURATION MODEL AND THE ENTROPY S OF THE CANONICAL ENSEMBLE

In this section we want to derive the relation

$$\Sigma = S - \Omega, \tag{A1}$$

where Σ is the entropy of the configuration model of simplicial complexes given by the logarithm of the number of graphs satisfying hard constraints on the generalized degree $k_{d,0}(r) = k_r$ for vertices $r = 1, \ldots, N$, i.e.,

$$\Sigma = \ln \left[\sum_G \prod_r \delta(k_r, k_{d,0}(r)) \right], \tag{A2}$$

S is the entropy of the canonical ensemble of simplicial complexes enforcing the expected generalized degree given by $\overline{k_{d,0}(r)} = k_r$ and Ω is the entropy of large deviation, which is the logarithm of the probability that in the canonical ensemble mentioned above, the generalized degree $k_{d,0}(r)$ take exactly the values $k_{d,0}(r) = k_r$. The entropy S is given by

$$S = -\sum_\alpha [p_\alpha \ln p_\alpha + (1 - p_\alpha) \ln (1 - p_\alpha)], \tag{A3}$$

while Ω is given by

$$\Omega = -\ln \left[\sum_G P(G) \prod_r \delta(k_r, k_{d,0}(r)) \right]. \tag{A4}$$

In order to derive Eq. (A1) we use the integral representation of the Kroenecker δ,

$$\delta(x, y) = \int_{-\pi}^{\pi} \frac{d\omega}{2\pi} e^{i\omega x - i\omega y}, \tag{A5}$$

obtaining for Σ,

$$\Sigma = \ln \left[\sum_G \prod_r \int_{-\pi}^{\pi} \frac{d\omega_r}{2\pi} e^{i\omega_r k_r - i\omega_r \sum_{\alpha | r \subset \alpha} a_\alpha} \right]. \tag{A6}$$

Summing over all simplicial complexes G is equivalent to summing over all adjacency tensors $\mathbf{a}$ of elements $a_\alpha = 0, 1$. Performing the sum we get

$$\Sigma = \ln \int_{-\pi}^{\pi} \left(\prod_r \frac{d\omega_r}{2\pi} \right) e^{i \sum_r \omega_r k_r} \prod_{\alpha \in \mathcal{Q}_d(N)} [1 + e^{-i \sum_{r \in \alpha} \omega_r}]$$

$$= \ln \int_{-\pi}^{\pi} \left(\prod_r \frac{d\omega_r}{2\pi} \right) e^{F(\boldsymbol{\omega}, \mathbf{k})}, \tag{A7}$$

where the function $F(\boldsymbol{\omega}, \mathbf{k})$ is given by

$$F(\boldsymbol{\omega}, \mathbf{k}) = i \sum_r \omega_r k_r + \sum_{\alpha \in \mathcal{Q}_d(N)} \ln[1 + e^{-i \sum_{r \in \alpha} \omega_r}]. \tag{A8}$$

We now apply a change of variables $\omega_r \to z_r$, where

$$\omega_r = z_r + \omega_r^\star, \tag{A9}$$

and where $\{\omega_r^\star\}$ indicates the solution to the equation

$$\frac{d}{d\omega_r} F(\boldsymbol{\omega}, \mathbf{k}) = 0. \tag{A10}$$

We note, that the above equations imply that the $\omega_r^\star$ are related to our choice of generalized degree sequence by

$$k_r = \sum_{\alpha | r \subset \alpha} \frac{e^{-i \sum_{m \subset \alpha} \omega_m^\star}}{1 + e^{-i \sum_{m \subset \alpha} \omega_m^\star}}. \tag{A11}$$

We can therefore identify $i \omega_r^\star$ with the parameters λ_r of the canonical ensemble in which we enforce that the expected generalized degree of the nodes $\overline{k_{d,0}(r)}$ take the same value of the hard constrained generalized degrees, i.e.,

$$\overline{k_{d,0}(r)} = k_r. \tag{A12}$$

In fact by setting $\lambda_r = i \omega_r^\star$, we get

$$p_\alpha = \frac{e^{-\sum_{m \subset \alpha} \lambda_m}}{1 + e^{-\sum_{m \subset \alpha} \lambda_m}} = \frac{e^{-i \sum_{m \subset \alpha} \omega_m^\star}}{1 + e^{-i \sum_{m \subset \alpha} \omega_m^\star}}, \tag{A13}$$

and Eq. (A11) reads

$$k_r = \overline{k_{d,0}(r)} = \sum_{\alpha | r \subset \alpha} p_\alpha. \tag{A14}$$

Writing $F(\boldsymbol{\omega}, \mathbf{k})$ in terms of our variable z_r given by Eq. (A9) we obtain

$$F(z, \boldsymbol{\omega}^\star, \mathbf{k}) = i \sum_r \omega_r^\star k_r$$
$$+ \sum_{\alpha \in \mathcal{Q}_d(N)} \ln[1 + e^{-i \sum_{r \in \alpha} \omega_r^\star}] + i \sum_r z_r k_r$$
$$+ \sum_{\alpha \in \mathcal{Q}_d(N)} \ln[1 - p_\alpha + p_\alpha e^{-i \sum_{r \in \alpha} z_r}], \tag{A15}$$

where p_α is given by Eq. (A13). We identify the first two terms of this expression with the entropy of the canonical ensemble S given by Eq. (20). In fact,

$$S = -\sum_\alpha [p_\alpha \ln p_\alpha + (1 - p_\alpha) \ln(1 - p_\alpha)]$$
$$= i \sum_r \omega_r^\star k_r + \sum_{\alpha \in \mathcal{Q}_d(N)} \ln[1 + e^{-i \sum_{r \in \alpha} \omega_r^\star}], \tag{A16}$$

where in the last expression we have substituted in the expression for p_α given by Eq. (A13). Finally, using Eq. (A16) and inserting the expression found for $F(z, \boldsymbol{\omega}^\star, \mathbf{k})$ [Eq. (A15)] back in to Eq. (A7) we obtain

$$\Sigma = \ln \left\{ e^S \int_{-\pi}^{\pi} \left(\prod_r \frac{dz_r}{2\pi} \right) e^{i \sum_r z_r k_r} \right.$$
$$\times \left. \prod_{\alpha \in \mathcal{Q}_d(N)} [1 - p_\alpha + p_\alpha e^{-i \sum_{r \in \alpha} z_r}] \right\}$$
$$= \ln \left[e^S \sum_{\{a_\alpha\}} \prod_\alpha p_\alpha^{a_\alpha} (1 - p_\alpha)^{1 - a_\alpha} \prod_r \delta \left(k_r, \sum_{\alpha' | r \subset \alpha'} a_{\alpha'} \right) \right].$$

Therefore, Σ can be written as

$$\Sigma = \ln \left[e^S \sum_G P(G) \prod_r \delta \left(k_r, \sum_{\alpha | r \subset \alpha} a_\alpha \right) \right], \tag{A17}$$

where $P(G)$ is given by Eq. (19). Using the definition of Ω given by Eq. (A4) it follows that

$$\Sigma = S - \Omega. \tag{A18}$$

APPENDIX B: DERIVATION OF THE EQ. (56) FOR Ω

In this Appendix we derive Eq. (56) for Ω in the presence of the structural cutoff. The quantity Ω indicates the logarithm of the probability that in a canonical ensemble of simplicial complexes enforcing the sequence of expected degree of the nodes $\{\overline{k_r} = k_r\}$ we observe a simplicial complex realization in which the sequence of the generalized degree of the nodes is exactly $\{k_r\}$. By indicating with $P(G)$ the probability of a simplicial complex in the canonical ensemble, we have

$$\Omega = -\ln \sum_G P(G) \prod_r \delta(k_r, k_{d,0}(r))$$
$$= -\ln \sum_G \prod_\alpha p_\alpha^{a_\alpha} (1 - p_\alpha)^{1 - a_\alpha} \prod_r \delta(k_r, k_{2,0}(r)), \tag{B1}$$

where, in presence of the structural cutoff, the probabilities p_α are given by Eq. (A13). In order to evaluate Ω, we use the integral representation of the Kroenecker δ,

$$\delta(x, y) = \int_{-\pi}^{\pi} \frac{d\omega}{2\pi} e^{i\omega x - i\omega y}, \tag{B2}$$

getting

$$\Omega = -\ln \sum_G \prod_\alpha p_\alpha^{a_\alpha} (1 - p_\alpha)^{1 - a_\alpha}$$
$$\times \prod_r \int_{-\pi}^{\pi} \frac{d\omega_r}{2\pi} e^{i\omega_r k_r - i\omega_r \sum_{\alpha' | r \subset \alpha'} a_{\alpha'}}$$
$$= -\ln \int_{-\pi}^{\pi} \prod_r \frac{d\omega_r}{2\pi} e^{\mathcal{G}[\{\omega_r\}]}, \tag{B3}$$

where

$$\mathcal{G}[\{\omega_r\}] = i \sum_i \omega_r k_r + \sum_\alpha \ln[1 + p_\alpha(e^{-i \sum_{r \subset \alpha} \omega_r} - 1)].$$

For an uncorrelated simplicial complex ensemble with structural cutoff and with p_α given by Eq. (26) and $p_\alpha \ll 1$ we can approximate $\mathcal{G}[\{\omega_r\}]$ as

$$\mathcal{G}[\{\omega_r\}] = i \sum_r \omega_r k_r + \sum_\alpha p_\alpha(e^{-i \sum_{r \subset \alpha} \omega_r} - 1). \tag{B4}$$

Using the explicit factorized expression for p_α in the presence of the structural cutoff [Eq. (26), we observe that we can write

$$\sum_\alpha p_\alpha(e^{-i \sum_{r \subset \alpha} \omega_r} - 1) = \frac{d!}{(d+1)!} \langle k \rangle N(\nu^{d+1} - 1), \tag{B5}$$

where

$$\nu = \sum_r \frac{k_r}{\langle k \rangle N} e^{-i\omega_r}. \tag{B6}$$

We now introduce the density

$$c(\omega | k) = \frac{1}{N_k} \sum_r \delta(\omega - \omega_r) \delta(k, k_r), \tag{B7}$$

where

$$N_k = N P_{d,0}(k) \tag{B8}$$

indicates the number of nodes with generalized degree of the nodes $k_r = k$ and $P_{d,0}(k)$ indicates the distribution of the generalized degree of the nodes. We can, therefore, express ν given by Eq. (B6) in terms of $c(\omega|k)$, obtaining

$$\nu = \sum_k \frac{k}{\langle k \rangle} P_{d,0}(k) \int d\omega e^{-i\omega} c(\omega|k). \tag{B9}$$

Using the δ functions,

$$\delta\left(c(\omega|k), \frac{1}{N_k} \sum_r \delta[\omega - \omega_r)\delta(k,k_r)]\right)$$

$$= \int_{-\pi}^{\pi} \frac{d\hat{c}(\omega|k)}{2\pi N_k} e^{i\hat{c}(\omega|k)[N_k c(\omega|k) - \sum_r \delta(\omega - \omega_r)\delta(k,k_r)]}, \tag{B10}$$

we can now express Ω as

$$\Omega = -\ln \int \mathcal{D}c(\omega|\hat{\lambda},k)\mathcal{D}\hat{c}(\omega|\hat{\lambda},k) e^{NF[c(\omega|k),\hat{c}(\omega|k)]},$$

where $F[c(\omega|k),\hat{c}(\omega|k)]$ is given by

$$F[c(\omega|k),\hat{c}(\omega|k)] = i \sum_k P_{d,0}(k) \int d\omega \hat{c}(\omega|k)c(\omega|k)$$

$$+ \frac{d!}{(d+1)!} \langle k \rangle (\nu^{d+1} - 1)$$

$$+ \sum_k P_{d,0}(k) \ln \int \frac{d\omega}{2\pi} e^{i\omega k - i\hat{c}(\omega|k)}. \tag{B11}$$

We evaluate the integral Eq. (B11) with the saddle point method. The saddle point equations read,

$$\frac{\partial F[c(\omega|k),\hat{c}(\omega|k)]}{\partial c(\omega|k)} = 0,$$

$$\frac{\partial F[c(\omega|k),\hat{c}(\omega|k)]}{\partial \hat{c}(\omega|k)} = 0,$$

which gives us

$$-i\hat{c}(\omega|k) = k\nu^d e^{-i\omega}, \tag{B12}$$

$$c(\omega|k) = \frac{\frac{1}{2\pi}e^{i\omega k - i\hat{c}(\omega|k)}}{\int \frac{d\omega}{2\pi} e^{i\omega k - i\hat{c}(\omega|k)}}. \tag{B13}$$

Using Eq. (B12) we observe that the integral appearing in Eq. (B13) can be expressed in terms of ν, obtaining

$$\int \frac{d\omega}{2\pi} e^{i\omega k - i\hat{c}(\omega|\hat{\lambda},k)} = \int \frac{d\omega}{2\pi} e^{i\omega k + k\nu^d e^{-i\omega}}$$

$$= \int \frac{d\omega}{2\pi} e^{i\omega k} \sum_h (\nu^d k)^h e^{-i\omega h} \frac{1}{h!}$$

$$= \frac{(\nu^d k)^k}{k!}. \tag{B14}$$

Substituting this result in to Eq. (B13), we get

$$c(\omega|k) = \frac{k!}{2\pi(\nu^d k)^k} e^{i\omega k - i\hat{c}(\omega|k)}, \frac{k!}{2\pi(\nu^d k)^k} e^{i\omega k + k\nu^d e^{-i\omega}}. \tag{B15}$$

Finally, we can substitute this expression in to the definition of ν given by Eq. (B9), obtaining

$$\nu = \sum_k \frac{k}{\langle k \rangle} P_{d,0}(k) \int d\omega e^{-i\omega} c(\omega|k)$$

$$= \sum_k \frac{k}{\langle k \rangle} P_{d,0}(k) \frac{k!}{(\nu^d k)^k} \int \frac{d\omega}{2\pi} e^{i\omega(k-1)+k\nu^d e^{-i\omega}}$$

$$= \sum_k \frac{k}{\langle k \rangle} P_{d,0}(k) \frac{k!}{(\nu^d k)^k} \int \frac{d\omega}{2\pi} e^{i\omega(k-1)} \sum_h (\nu^d k)^h e^{-i\omega h} \frac{1}{h!}$$

$$= \sum_k \frac{k}{\langle k \rangle} P_{d,0}(k) \frac{k!}{(\nu^d k)^k} \frac{(\nu^d k)^{k-1}}{(k-1)!} = \nu^{-d}. \tag{B16}$$

Therefore, ν is the solution of the equation $\nu = \nu^{-d}$, and so we have

$$\nu = 1. \tag{B17}$$

Using this result, and Eq. (B14) it is immediate to show that the value of the functional $F[c(\omega|k),\hat{c}(\omega|k)]$ [Eq. (B11)] at the saddle point is given by

$$F[c(\omega|k),\hat{c}(\omega|k)] = i \sum_k P_{d,0}(k) \int d\omega \hat{c}(\omega|k)c(\omega|k)$$

$$+ \sum_k P_{d,0}(k) \ln \left[\frac{k^k}{k!} \right]. \tag{B18}$$

Proceeding as in Eq. (B16), it can be easily shown that

$$i \sum_k P_{d,0}(k) \int d\omega \hat{c}(\omega|k)c(\omega|k)$$

$$= -\sum_k P_{d,0}(k)k \frac{k!}{k^k} \int d\omega 2\pi e^{i\omega(k-1)+ke^{-i\omega}}$$

$$= -\langle k \rangle = -\sum_k P_{d,0}(k)k. \tag{B19}$$

Finally, evaluating the integral Eq. (B11) at the saddle point we obtain the simple expression for Ω given by

$$\Omega = -\ln \left[N \sum_k P_{d,0}(k) \ln \left(\frac{k^k}{k!} e^{-k} \right) \right]$$

$$= -\ln \left[\sum_r \ln(\pi_{k_r}(k_r)) \right], \tag{B20}$$

where $\pi_{k_r}(k_r)$ indicated the Poisson distribution with average k_r calculated at k_r; i.e.,

$$\pi_{k_r}(k_r) = \frac{k_r^{k_r}}{k_r!} e^{-k_r}. \tag{B21}$$

[1] G. Bianconi, Europhys. Lett. **111**, 56001 (2015).

[2] S. Boccaletti, G. Bianconi, R. Criado, C. I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, and M. Zanin, Phys. Rep. **544**, 1 (2014).

[3] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, J. Complex Netw. **2**, 203 (2014).

[4] G. Ghoshal, V. Zlatic, G. Caldarelli, and M. E. J. Newman, Phys. Rev. E **79**, 066118 (2009).

[5] V. Zlatic, G. Ghoshal, and G. Caldarelli, Phys. Rev. E **80**, 036118 (2009).

[6] C. Giusti, R. Ghrist, and D. S. Bassett, arXiv:1601.01704 (2016).

[7] P. Dotko *et al.*, arXiv:1601.01580 (2016).

[8] M. E. J. Newman and J. Park, Phys. Rev. E **68**, 036122 (2003).

[9] E. Agliari, A. Annibale, A. Barra, A. C. C. Coolen, and D. Tantari, J. Phys. A **46**, 415003 (2013).

[10] R. Schifanella, A. Barrat, C. Cattuto, B. Markines, and F. Menczer, in *Proceedings of the Third ACM International Conference on Web Search and Data Mining* (ACM Press, New York, 2010), pp. 271–280.

[11] M. Szomszor, C. Cattuto, H. Alani, K. OHara, A. Baldassarri, V. Loreto, and V. D. P. Servedio, in *4th European Semantic Web Conference, Bridging the Gap between Semantic Web and Web 2.0, Innsbruck, Austria, June 3–7* (2007).

[12] Z. Wu, G. Menichetti, C. Rahmede, and G. Bianconi, Sci. Rep. **5**, 10073 (2015).

[13] G. Bianconi, C. Rahmede, and Z. Wu, Phys. Rev. E **92**, 022815 (2015).

[14] G. Bianconi and C. Rahmede, Sci. Rep. **5**, 13979 (2015).

[15] G. Bianconi and C. Rahmede, Phys. Rev. E **93**, 032315 (2016).

[16] O. Eisenberg, D. Krioukov, and K. Zuev, J. Phys. A **48**, 465002 (2015).

[17] G. Petri, M. Scolamiero, I. Donato, and F. Vaccarino, PLoS ONE **8**, e66506 (2013).

[18] G. Petri, P. Expert, F. Turkheimer, R. Carhart-Harris, D. Nutt, P. J. Hellyer, and F. Vaccarino, J. Roy. Soc. Interface **11**, 20140873 (2014).

[19] L. Lü, M. Medo, C. H. Yeung, Y.-C. Zhang, Z.-K. Zhang, and T. Zhou, Phys. Rep. **519**, 1 (2012).

[20] D. Taylor, F. Klimm, H. A. Harrington, M. Kramar, K. Mischaikow, M. A. Porter, and P. J. Mucha, Nat. Commun. **6**, 7723 (2015).

[21] C. V. Cannistraci, G. Alanis-Lobato, and T. Ravasi, Sci. Rep. **3**, 1613 (2013).

[22] M. A. Serrano, D. Krioukov, and M. Boguná, Phys. Rev. Lett. **100**, 078701 (2008).

[23] F. Papadopoulos, M. Kitsak, M. A. Serrano, M. Boguñá, and D. Krioukov, Nature **489**, 537 (2012).

[24] J. Ambjorn, J. Jurkiewicz, and R. Loll, Phys. Rev. D **72**, 064014 (2005).

[25] J. Ambjorn, J. Jurkiewicz, and R. Loll, Phys. Rev. Lett. **93**, 131301 (2004).

[26] D. Oriti, Rep. Prog. Phys. **64**, 1703 (2001).

[27] S. Gielen, D. Oriti, and L. Sindoni, Phys. Rev. Lett. **111**, 031301 (2013).

[28] J. Park and M. E. J. Newman, Phys. Rev. E **70**, 066117 (2004).

[29] K. Anand, and G. Bianconi, Phys. Rev. E **80**, 045102 (2009).

[30] C. I. Del Genio, H. Kim, Z. Toroczkai, and K. E. Bassler, PLoS ONE **5**, e10012 (2010).

[31] H. Kim, C. I. Del Genio, K. E. Bassler, and Z. Toroczkai, New J. Phys. **14**, 023012 (2012).

[32] G. Bianconi, Europhys. Lett. **81**, 28005 (2008).

[33] G. Bianconi, Phys. Rev. E **79**, 036114 (2009).

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

[35] K. Anand and G. Bianconi, Phys. Rev. E **82**, 011116 (2010).

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

[37] O. Sagarra, C. J. Pérez Vicente, and A. Díaz-Guilera, Phys. Rev. E **88**, 062806 (2013).

[38] O. Sagarra, C. J. Pérez Vicente, and A. Díaz-Guilera, Phys. Rev. E **92**, 052816 (2015).

[39] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).

[40] A. Costa and M. Farber, arXiv:1412.5805 (2014).

[41] D. Cohen, A. Costa, M. Farber, and T. Kappeler, Discrete Comput. Geom. **47**, 117 (2012).

[42] M. Kahle, AMS Contemp. Math **620**, 201 (2014).

[43] G. Bianconi, P. Pin, and M. Marsili, Proc. Natl. Acad. Sci. USA **106**, 11433 (2009).

[44] J. Iacovacci, Z. Wu, and G. Bianconi, Phys. Rev. E **92**, 042806 (2015).

[45] E. A. Bender and E. R. Canfield, J. Comb. Theor., Ser. A **24**, 296 (1978).

[46] M. Boguná, R. Pastor-Satorras, and A. Vespignani, Eur. Phys. J. B **38**, 205 (2004).

[47] P. Erdös and T. Gallai, Mat. Lapok **11**, 477 (1960).

[48] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.93.062311 for the codes generating random simplicial complexes in $d = 1$, $d = 2$, and $d = 3$.